Essays in Empirical Microeconomics

Dissertation
submitted to the
Faculty of Business, Economics and Informatics
of the University of Zurich

to obtain the degree of
Doktorin der Wirtschaftswissenschaften, Dr. oec.
(corresponds to Doctor of Philosophy, PhD)

presented by

Janina Nemitz
from Germany

approved in July 2018 at the request of

Prof. Dr. Rainer Winkelmann
Prof. Dr. Pietro Biroli

The Faculty of Business, Economics and Informatics of the University of Zurich hereby authorizes the printing of this dissertation, without indicating an opinion of the views expressed in the work.

Zurich, July 18^{th} 2018

Chairman of the Doctoral Board: Prof. Dr. Steven Ongena

Acknowledgments

Undertaking my PhD studies would not have been possible without substantial scientific support and guidance that I received from many people. Beyond acknowledging comments and suggestions on the individual chapters, which are given therein, I wish to thank a few persons that especially supported me during the process of writing this dissertation:

First and foremost, I wish to express my gratitude to my supervisor Rainer Winkelmann and my co-supervisor Pietro Biroli. I am grateful for their profound guidance, their dedicated mentoring, and the support that I experienced from them over the last five years. I cannot thank them enough for their patience and assistance.

I am also more than grateful to my co-authors Wanda Mimra and Christian Waibel who have been my main collaborators at the fourth chapter of this dissertation. I very much enjoyed the interactions and discussions with them, and I tremendously benefited from their experience and advice, not only academically but also personally.

Further, I thank my mentor Johannes Kunz for his support and inspiring conversations. I also wish to thank Luca Tonizzo and Emina Boudemagh for their excellent research assistance. Moreover, I wish to acknowledge organizational assistance by Mirjam Britschgi, Christine Le Grand, Bettina Petralli, and Maura Wyler-Zerboni.

I also thank my friends at the University of Zurich, in particular, Kathrin Degreiff, Regina Hammerschmid, and Harald Mayr for making my time at the University of Zurich very pleasant and unforgettable.

Finally, I wish to express my gratitude to my parents and siblings. They always infinitely support me, encourage me, and fully believe in what I do. I thank my partner Johannes, whom I value more than anyone else for his patience, infinite understanding, and ongoing support.

Janina Nemitz

Zurich, March 2018

Contents

List of Tables

List of Figures

List of Abbreviations

ADSP All-Day School Program

ECU Experimental Currency Unit

FP Full Pooling

GINA Genetic Information Nondiscrimination Act

GPO Genetic Pooling Only

IZBB Investition Zukunft Bildung und Betreuung

OECD Organisation for Economic Cooperation and Development

QALY Quality-Adjusted Life Year

SOEP Socio-Economic Panel

2SLS Two-Stage Least-Squares

WTP Willingness To Pay

Chapter 1

Dissertation Overview

Throughout the last decades, human longevity has substantially improved all over the world. As a consequence, the world's population aged 60 and over increased from 607.1 million in 2000 to 900.9 million in 2015. By 2050, it is expected to more than double its size, accounting for more than 20% of the overall population in the world, which corresponds to a 220%-increase relative to the proportion of elderly in 2000 (United Nations 2015). This global trend has far-reaching economic and socio-political implications for societies, including a decline in the share of the working-age population, increased health care costs, and unsustainable pension commitments, all of which may offset the potential benefits of living longer. Hence, a major challenge for aging societies is to address these issues. Individual, family, and societal resources might need to be reallocated to support growing lifespans and a good quality of life. Policymakers must account for new priorities when allocating their scare resources, in order to assist the growing population of elderly, while still taking into account the problems of younger populations and limiting the redistributive burden for those at working ages.

The chapters in this cumulative dissertation consider three important topics in the field of empirical microeconomics that relate to current issues arising in aging societies. Chapter 2, titled *Increasing Life Expectancy and Life Satisfaction: Is There a Catch to Living Longer?*, addresses a pressing problem in the allocation of scarce resources, namely the tradeoff between lengthening human life and enhancing the quality of human life. It challenges the prominent view that more (in this case longer lives) is always better and acknowledges that longer lives might come at the price of lower quality and life satisfaction in old age, possibly extending the lifetime spent in states of dissatisfaction. By investigat-

ing how end-of-life life satisfaction and cumulative life satisfaction beyond the age of 60 changed with improved longevity over time, this chapter makes a seminal contribution to the literature (e.g., Crimmins and Beltrán-Sánchez (2011) and Chatterji et al. (2015) for two recent surveys), whose main focus has hitherto been on health-related outcomes. It shows that both end-of-life life satisfaction and the proportion of satisfied lifetime to total remaining lifetime decreased for West German elderly between 1985 and 2010, thereby raising doubts over the benefits of resources allocated to life-extending policies. Moreover, by uncovering that health and social isolation strongly contributed to the decline in life satisfaction over time, it sheds light on which types of quality-of-life improving policy options would be most effective for policymakers to undertake in the future.

Chapters 3 and 4 both investigate policies that have the potential to limit the redistributive burden for younger populations in aging societies. Chapter 3, titled *The Effect of All-Day Primary School Programs on Maternal Labor Supply*, considers a policy that aimed at activating a large unused source of labor force in Germany, namely that of mothers with primary school-aged children, by expanding low-cost childcare supply. The activation of unused skill potentials is a challenging task in aging societies to decrease the old-age dependency ratio and to secure pension commitments in the short run until a gradual increase in the statutory pension age is fully implemented. In 2003, the German federal government launched the public investment program "Future Education and Care" (IZBB) to expand all-day primary school programs (ADSPs) in Germany. Lengthening the time spent in primary schools beyond half day at essentially zero costs for parents, ADSPs significantly reduce the opportunity costs of maternal employment and, thus, are expected to increase maternal labor supply. To analyze the effect of the voluntary ADSPs on maternal labor supply, bivariate probit models are estimated. Exogenous variation in the allocation of IZBB investments across time and counties is used to identify these models.

This chapter adds to the vast literature that analyzes the effect of childcare programs on maternal labor supply (e.g., Gelbach 2002, Baker et al. 2008, Cascio 2009, Fitzpatrick 2010, Havnes and Mogstad 2011, and Bauernschuster and Schlotter 2015), whose main focus has been on childcare programs for preschool-aged children. It demonstrates that ADSPs have a large effect on maternal labor supply at the extensive margin, but may have no effect on maternal labor supply at the intensive margin (full-time vs. part-time

employment) if they remain incompatible with full-time working schedules. Thus, while improving unfavorable old-age dependency ratios in the short run, ADSPs with limited operating hours are as yet unable to activate the full labor force potential of mothers with primary school-aged children in Germany. In particular, they fail to increase maternal labor supply at the intensive margin; mothers who resume employment often only engage in part-time employment. Also in terms of additional taxes collected to pay for policies that benefit the elderly, these programs may yield fewer benefits than expected. Chapter 3 shows for Germany that labor supply responses at the extensive margin are concentrated among mothers with at most a vocational degree, i.e. among those mothers who are paid lower wages, on average. Overall, this chapter stresses the importance of voluntary ADSPs in increasing labor supply of mothers with primary school-aged children, but it also discusses some unintended consequences that may arise if these programs clash with full-time working schedules (e.g., increase of the gender gap in working hours and wages), eventually suggesting an extension of these programs to full day.

Chapter 4, titled *Voluntary Pooling of Genetic Risk: A Health Insurance Experiment* and jointly written with Wanda Mimra and Christian Waibel, considers a policy that aims at reducing the burden of cross-subsidization from young, healthy populations to old, unhealthy populations in statutory health insurance by allowing health insurance providers to individually price behavioral health risks. Separating out behavioral health risks, which are subject to individual manipulation, eliminates free-riding incentives in group health insurance and may thereby reduce excessive health care costs, particularly in old age, where behavioral health risks accumulate. Beyond that, it may increase the willingness to pool on health risks that are uncontrollable by individuals (e.g., higher genetic risk in old age), thereby increasing voluntary participation in group insurance among people with low health risks. This would lead to lower health insurance premiums in dual systems (i.e., in systems where social and private health insurance coexist) or increase political support in systems with obligatory social health insurance only. Scientific and technological advances in detecting, estimating, and monitoring health risks that allow for better tailoring of health plans to individual health risk profiles render such a policy feasible.

To assess the consequences of separating out and individually pricing behavioral health risk in voluntary group health insurance, chapter 4 uses an incentivized laboratory ex-

periment. In the experiment, subjects' overall health risk has an assigned, uncontrollable genetic risk part and a behavioral risk part, which can be reduced by costly effort. The experimental variation either includes behavioral risk in the pooling of a group insurance scheme or separates it out. Outside options are an individual risk-based insurance, i.e. a health insurance which individually prices genetic and behavioral health risks, and no insurance. This chapter shows that people exhibit social preferences for pooling in health insurance. Due to both large heterogeneity in social preferences for pooling across subjects and the dynamics of the willingness to pay for group insurance in the different experimental markets, only a low level of actual genetic risk pooling is observed across the two experimental conditions, however. Although there is a tendency towards more voluntary pooling on genetic risk if group health insurance individually prices behavioral health risk, chapter 4 suggests that mandatory pooling might be needed if, under the veil of ignorance, a society nevertheless wishes to pool certain forms of heterogeneous risk (e.g., higher non-modifiable genetic risk in old age). In aging societies such a mandate is, however, more likely to be broadly accepted among young, healthy populations if group insurance individually prices behavioral risks, thereby reducing the burden of cross-subsidization stemming from health risks that are controllable by individuals and disproportionately high in old age.

Beyond this thematic linkage, there is also a central methodological feature underlying all chapters: All chapters strive towards solid evidence and correct inference. In social sciences, this generally requires carefully addressing a number of issues, including non-random treatment assignment and selection into treatment, reverse causality, and omitted variables. Given the large toolbox of econometric methods, the best feasible empirical strategy has to be selected. This strategy strongly depends on the data at hand and, thus, the most appropriate data set should be chosen. Empirical strategies often rely on some key assumptions. Therefore, a discussion of the key assumptions and, if possible, the provision of supportive evidence for their validity is an integral part of any empirical work. Another important issue is that of correct inference. In most empirical applications the standard assumption of independently and identically distributed errors is not fulfilled. In panel data, for example, standard errors need to be clustered to account for within-cluster correlation, and additional adjustments are required if the number of clusters is small. Finally, evidence should be robust. In view of the numerous decisions

that researchers make throughout data preparation and estimation, it is important to show that results do not depend on these decisions.

All chapters in this dissertation take this seriously. Chapter 2, for example, examines and rules out a large set of explanations other than improved longevity for the decline in life satisfaction over time. It discusses problems such as omitted variables, reverse causality, and attrition. Moreover, it makes use of the most appropriate data set available. The German Socio-Economic Panel (SOEP) is unique in that it combines three important features: First, it provides information on overall life satisfaction. Second, it does so for an extensive period of time, which coincides with substantial increases in life expectancy. And third, it tracks respondents over time up until death, thereby allowing for a novel approach that gives rise to more detailed results whose sensitivity can be more carefully explored than in earlier work. This approach also permits testing an important theory, namely that of terminal life satisfaction decline, and corroborates this theory by showing that it also holds in aging societies across time. In addition, correct inference is an important issue that surfaces in this chapter. Given the panel structure of the data, standard errors are clustered. For the second approach taken in this chapter, new standard errors are derived based on the delta method because, given the SOEP data, they differ from those previously used in the literature for repeated cross-sectional data.

The central issue that is addressed in chapter 3 is that of non-random assignment to and selection into treatment. Non-random assignment to treatment results from the fact that in Germany schools with ADSP are allowed to reject students if capacity constraints of an ADSP are reached, in which priority is often given to children of mothers who are more prone to employment (selectivity of schools with ADSP). Selection into treatment results from the fact that in Germany ADSPs are made available to parents on a voluntary basis (selection into ADSPs). In order to deal with both types of selection, this chapter uses an instrumental variables strategy and estimates bivariate probit models. The key identifying assumption is the exclusionary restriction. Thus, a large part of chapter 3 is devoted to the discussion of this assumption and several tests are performed to provide some suggestive evidence for the validity of this assumption. Another important issue that surfaces in this chapter is that of robustness of results. A large battery of robustness checks is employed to test the sensitivity of results to alternative estimation methods, coding decisions made throughout the data preparation process, and alternative model

specifications. The role of maternal work preferences, which are omitted in most applications due to the lack of data, is explored, and the sensitivity of results to potentially endogenous regressors (e.g., non-wife household income) is investigated.

Unlike chapter 3, chapter 4 makes use of the gold standard and randomly assigns subjects to treatment. Although random assignment of treatment solves many issues present in empirical work, it does not mean that inference based on experiments is free of errors. Apart from the fact that random assignment may fail in small samples leading to unbalanced characteristics of treatment and control groups, even in a controlled environment such as the laboratory experimental evidence can be misleading if, for example, key parameters are not carefully selected. Therefore, a clean and simple experimental design is key for the superiority of experiments. The incentivized laboratory experiment in chapter 4 makes use of several important design features, three of which are highlighted here: The outside option individual risk-based insurance is used to allow for a clear identification of social preferences. The assumption of additive separability of genetic and behavioral health risk is applied to allow disentangling social preferences for pooling on genetic risk from those for pooling on behavioral risk. Finally, an important feature employed to make the health prevention decision more salient is that of tying the preventive effort decision inside the laboratory to the probability of winning a voucher for a health preventative measure outside of the laboratory. While experiments are straightforward to evaluate ex post, there are still some important methodological aspects that surface in this chapter. For example, non-parametric tests are used to account for the small number of independent observations, while the wild cluster bootstrap t-procedure suggested in Cameron et al. (2008) is applied to deal with interrelated decisions of subjects in the presence of few clusters.

References

Baker, M., J. Gruber, and K. Milligan (2008): "Universal Child Care, Maternal Labor Supply, and Family Well-Being," *Journal of Political Economy*, 116, 709–745.

Bauernschuster, S. and M. Schlotter (2015): "Public Child Care and Mothers' Labor Supply - Evidence from Two Quasi-Experiments," *Journal of Public Economics*, 123, 1–16.

Cameron, A. C., J. B. Gelbach, and D. L. Miller (2008): "Bootstrap-Based Improvements for Inference with Clustered Errors," *Review of Economics and Statistics*, 90, 414–427.

Cascio, E. U. (2009): "Maternal Labor Supply and the Introduction of Kindergartens into American Public Schools," *Journal of Human Resources*, 44, 140–170.

Chatterji, S., J. Byles, D. Cutler, T. Seeman, and E. Verdes (2015): "Health, Functioning, and Disability in Older Adults – Present Status and Future Implications," *Lancet*, 385, 563–575.

Crimmins, E. M. and H. Beltrán-Sánchez (2011): "Mortality and Morbidity Trends: Is There Compression of Morbidity?" *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 66B, 75–86.

Fitzpatrick, M. (2010): "Preschoolers Enrolled and Mothers at Work? The Effects of Universal Prekindergarten," *Journal of Labor Economics*, 28, 51–85.

Gelbach, J. B. (2002): "Public Schooling for Young Children and Maternal Labor Supply," *American Economic Review*, 92, 307–322.

Havnes, T. and M. Mogstad (2011): "Money for Nothing? Universal Child Care and Maternal Employment," *Journal of Public Economics*, 95, 1455–1465.

United Nations, Department of Economic and Social Affairs, Population Division (United Nations) (2015): "World Population Ageing 2015 (ST/ESA/SER.A/390)," `www.un.org/en/development/desa/population/publications/pdf/ageing/WPA2015_Report.pdf`. Accessed on 02 March 2018.

Chapter 2

Increasing Life Expectancy and Life Satisfaction: Is There a Catch to Living Longer?

Abstract: Human longevity is rising rapidly all over the world, but are longer lives more satisfied lives? This study suggests that the answer might be no. Despite living longer, people did not become better off in terms of overall life satisfaction because there were substantial losses of life satisfaction in old age. When compared to 1985, in 2010 West German elderly were, on average, much less satisfied throughout their final period of life. Moreover, they were expected to spend a larger proportion of their remaining lifetime in states of dissatisfaction. Two important mechanisms that contribute to this decline in satisfaction are health and social isolation. Using a broad variety of sensitivity tests, I show that these results are robust to a large set of alternative explanations.

JEL classification: I10, I31, J11
Keywords: Happy life expectancy, longevity, life satisfaction, time-to-death, Sullivan's method

Acknowledgements: I would like to thank Pieter Bakx, Pietro Biroli, Friedrich Breyer, Kristina Hess, Johannes Kunz, Steven Lehrer, Harald Mayr, Andrew Oswald, Florian Schaffner, Hannes Schwandt, Rainer Winkelmann, seminar participants in Zams, Zurich, and Konstanz, and participants at the 12th International German Socio-Economic Panel User Conference in Berlin, the 2017 Annual Congress of the Swiss Society of Economics and Statistics in Lausanne, the 31st Conference of the European Society of Population Economics in Glasgow, the 2017 World Congress of the International Health Economics Association in Boston, the 4th European Health Economics Association PhD Student-Supervisor and Early Career Researcher Conference in Lausanne, and the 3rd Swiss Health Economics Workshop in Lausanne for comments and suggestions.

2.1 Introduction

Throughout the last decades, human life expectancy has strongly increased in OECD countries. As illustrated by Figure 2.1, the rise in life expectancy holds true for both men and women, at various ages, and across countries. To date, there are little signs of a slowdown of this positive trend, suggesting that there is scope for further improvements.[1] But are longer lives more satisfied lives? This question is vital for individuals and public policymakers. Longer lives might come at the price of lower quality and life satisfaction in old age. They will be less valuable to people if the additional life years are spent in dissatisfaction.[2] These issues are pivotal in the context of private and public decision-making, where decision makers face the tradeoff between lengthening human life and enhancing the quality of life.

Yet, little is known about how life satisfaction in old age has changed with improved longevity over time. Previous literature has almost exclusively focused on health-related measures (e.g., Crimmins et al. 2009, Cutler et al. 2014)[3], notwithstanding that health is only one determinant of life satisfaction (Easterlin 2002, 2003) and people partially adapt to poor states of health (Oswald and Powdthavee 2008, McNamee and Mendolia 2014). Evidence on life satisfaction and related concepts of well-being is limited to two studies that connect improvements in longevity to changes in well-being in old age by estimating time trends of cumulative life satisfaction (Perenboom et al. 2004, Yang 2008). The reliance on cumulative life satisfaction alone is, however, very restrictive because little is revealed about the underlying changes in life satisfaction at specific ages or in the last years of life, which are critical for end-of-life decision making. Moreover, without any additional analyses it is impossible to assess which factors other than improved longevity contribute to the changing life satisfaction patterns over time.

[1] One exception are the United States, with decreasing life expectancy for some disadvantaged subgroups of the population (e.g., Olshansky et al. 2012, Chetty et al. 2016). See also Case and Deaton (2015), for evidence on increasing mortality among middle-aged white Americans.

[2] Note that in countries with restrictive euthanasia laws longer lives might even backfire as there exist states of life which are considered worse than death (e.g., Ditto et al. 1996, Rubin et al. 2016). If improvements in longevity are such that people reach these states more often or spend more time in these states, longer lives might lower people's welfare.

[3] See Crimmins and Beltrán-Sánchez (2011), Jagger and Robine (2011), Chatterji et al. (2015), and Lindgren (2016) for four recent reviews. Studies in this field analyze time trends of either disease prevalence among the elderly or summary measures such as quality- or disability-adjusted life years or healthy life expectancy to test the three main hypotheses in the literature (Gruenberg 1977, Olshansky et al. 1991: expansion of morbidity; Fries 1980: compression of morbidity; Manton 1982: dynamic equilibrium). Results are mixed and strongly depend on the health indicator.

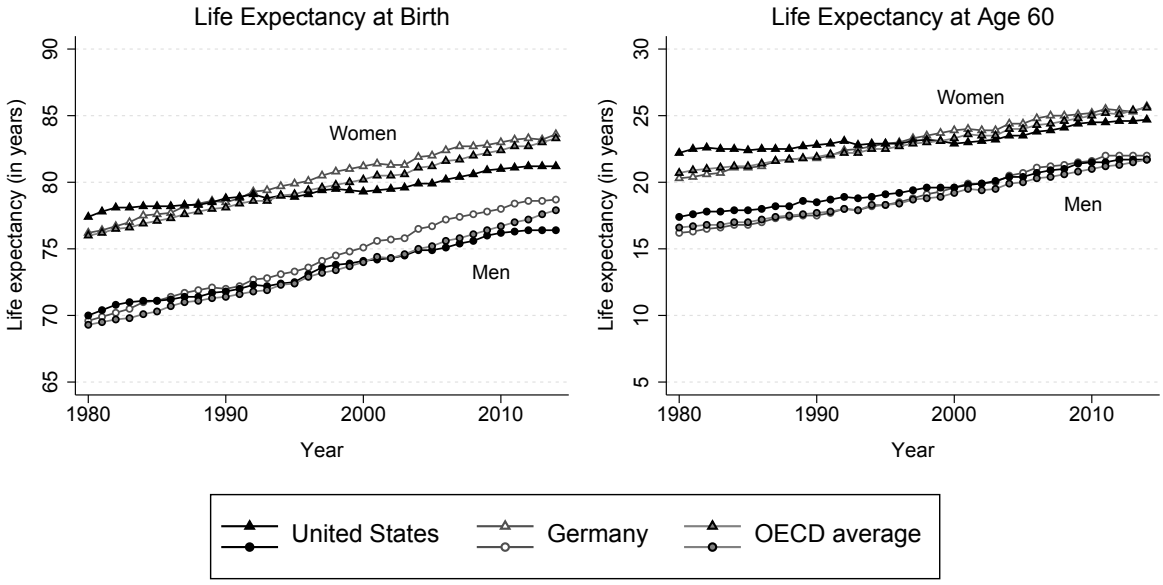

Figure 2.1. Life Expectancy at Birth and at Age 60 by Gender, Year, and Country

Source: OECD (2016), own representation.

In this study, I use a combination of two approaches paired with a broad set of sensitivity checks to investigate how life satisfaction in old age has changed with improved longevity in West Germany since 1985. I use data of the longest running household panel with continuous information on overall life satisfaction, the German Socio-Economic Panel (SOEP), and I analyze changes in both end-of-life life satisfaction (time-to-death approach) and expected cumulative life satisfaction beyond the age of 60 (life-expectancy approach). Using data of roughly 2,500 West German SOEP respondents who were within five years of death between 1985 and 2011, I estimate time trends of average life satisfaction by time to death. As life satisfaction is relatively stable in old age and only strongly declines in the last years of life (e.g., Gerstorf et al. 2008a,b, 2010), this approach is useful to detect changes in both the onset and the rate of terminal life satisfaction decline. Yet, this approach does not allow weighting gains in satisfied lifetime against losses in end-of-life life satisfaction.

Therefore, as a complementary approach, I estimate time trends of satisfied life expectancy at age 60 based on Sullivan's method (Sullivan 1971). Satisfied life expectancy at age 60 is a summary measure that collapses age-specific mortality and satisfaction prevalence rates observed in a given year into a single number. It provides information

on how long a cohort of 60-year-old survivors can expect to live a satisfied life under the assumption that age-specific mortality and satisfaction prevalence rates in that given year remain constant in the future. In order to weight increases in satisfied lifetime against increases in dissatisfied lifetime, I compare the proportion of satisfied life expectancy on total life expectancy at age 60 over time. Assuming that people prefer a high proportion of satisfied lifetime, the latter is used to assess whether the overall quality of life beyond the age of 60, on average, improved with increased longevity.

The results suggest that life satisfaction in old age strongly declined with improved longevity in West Germany. Between 1985 and 2010, West German elderly became, on average, much more dissatisfied throughout their last years of life. In any of the five years before death, average life satisfaction scores were roughly half a standard deviation lower in 2010. In addition, the slope of terminal life satisfaction decline was smaller. Both of these findings are consistent with an extension of the dissatisfied period at the end of people's life. Thus, it is not very surprising that dissatisfied life expectancy at age 60 increased with improved longevity between 1985 and 2010. I also find an increase in satisfied life expectancy at age 60 over time. Yet, this increase was too small to compensate for the increase in dissatisfied life expectancy. Therefore, in 2010, 60-year-old survivors were expected to spend a larger proportion of their remaining lifetime in states of dissatisfaction. I show that compositional changes in the population (e.g., income, education), cohort effects, time-in-panel effects, a simple aging-unrelated time trend, endogenous onset of disease and terminal life satisfaction decline, decreasing sample selectivity, and attrition cannot explain the decline in life satisfaction.

I explore potential mechanisms to understand how increased longevity contributed to the deteriorations of life satisfaction in old age. I find that both health and social isolation are important mechanisms. Several health indicators (e.g., severe disability, number of hospitalizations) suggest a deterioration of the end-of-life health status over time. The increase of legally attested disability had the most detrimental impact on end-of-life life satisfaction. Social isolation is measured by the frequency of mutual visits with family and friends. I find that both types of visits became less frequent with improved longevity. Individual-level life satisfaction regressions corroborate these descriptive findings and show that both health and social isolation explain a large fraction of the decline in end-of-life life satisfaction over time.

In terms of the life-expectancy approach, this study is closely related to Perenboom et al. (2004) and Yang (2008) who also estimated time trends of satisfied life expectancy within a country over time.[4] They show for the Netherlands and the United States that the expected lifetime in states of satisfaction increased in both absolute (number of years) and relative terms (proportion of life) in the 1980s and 1990s. Compared to this study, the focus of their studies is on different countries and earlier time periods. They also use different measures of well-being, which explains the discrepancy in findings. Perenboom et al. (2004) uses items of the negative affect balance scale, whereas Yang (2008) uses happiness – a concept which is more closely related to life satisfaction. In the latter study, however, life satisfaction is only measured on a three-point scale. In this study, I use a much more detailed life satisfaction measure, which allows me to test the sensitivity of results to alternative cutoff values when distinguishing between three states of satisfaction. I show that I can replicate earlier studies' findings if using an equal point split classification, but that this classification hides important patterns across time as it pools states of satisfaction and dissatisfaction in the intermediate category.

This study contributes to the existing literature in three important ways. First, this study is the first to analyze how life satisfaction in the final period of life has changed with improved longevity over time. It documents that the terminal life satisfaction decline also holds across time, thereby extending upon the small but growing literature on end-of-life life satisfaction (e.g., Gerstorf et al. 2008a,b, 2010, Palgi et al. 2010, Berg et al. 2011).[5] Rather than evaluating one particular medical innovation (e.g., drug eluding stents), this study evaluates the sum of all technological innovations up to a certain point in time. This comprehensive approach can account for spillovers across technologies and is important if new technologies affect both quality and length of life. Second, contrary to previous studies that exploit variation in longevity over time, this study carefully explores the role

[4]There are three additional studies which empirically assess satisfied life expectancy (Kunst et al. 1994, Veenhoven 1996, Yang and Waliji 2010). The former two studies compare satisfied life expectancy across countries and show that there exist large differences across countries (e.g., higher scores in rich Western European nations, Australia, and the United States). Yang and Waliji (2010) use a variant of the multistate life table method to analyze differences in satisfied life expectancy across social groups in the United States. They find lower satisfied life expectancy scores for male, black, and less educated Americans.

[5]This study also relates to the numerous studies that analyze the relationship between age and life satisfaction, in general. These studies often, but not always, show that the relationship between age and life satisfaction is u-shaped with a strong dip of life satisfaction in mid age, and possibly another downturn in old age (e.g., Blanchflower and Oswald 2004, 2008, 2017, Frijters and Beatton 2012, Wunder et al. 2013). Yet, the downturn in old age almost vanishes upon controlling for time to death (e.g., Gerstorf et al. 2008a,b, 2010).

of explanations other than improved longevity for the decline in life satisfaction. Therefore, I can exclude a large set of alternative explanations. Third, this study furthers our understanding of successful aging by shedding light on two important mechanisms: health and social isolation. Although these mechanisms are not new to the literature (e.g., Oswald and Powdthavee (2008) for the former and Helliwell (2003, 2006) for the latter), this study is the first to show that they also play an important role in aging societies across time.

The remainder of this paper is structured as follows. Section 2.2 discusses the framework that motivates the two empirical approaches of this study. Section 2.3 presents the data. Section 2.4 describes the time-to-death approach in detail and presents its results. Section 2.5 describes the life-expectancy approach in detail and presents the results of this approach. Section 2.6 discusses two potential mechanisms and shows some evidence for these mechanisms. Section 2.7 concludes and provides some policy implications.

2.2 A Framework

To evaluate the impact of improved longevity, it is necessary to compare the overall quality and goodness of two representative lives that differ with respect to their length. In economics, the quality and goodness of life is measured by lifetime utility. Lifetime utility captures the idea that people attach value to both length and quality of life, but also accounts for the fact that the value of a longer life strongly depends on its quality. It is often considered the primary target of public policy. As life satisfaction may be a good proxy for utility (Benjamin et al. 2012, 2014, Fleurbaey and Schwandt 2015), a measure of overall life satisfaction, which is particularly attractive to economists, is naturally obtained by replacing contemporaneous utility scores in the lifetime utility function with their corresponding reported life satisfaction scores. Assuming that people attach equal weight to each year of life (i.e. no discounting), overall life satisfaction of the representative agent is given by

$$TLS = \sum_{a=0}^{A} \left(LS_a - LS_d \right), \tag{2.1}$$

where $A \in \mathbb{R}_+$ is the age in the last year of life, $LS_a \in \left[\underline{LS}, \overline{LS}\right]$ is the life satisfaction score at age a, and $LS_d \in \left[\underline{LS}, \overline{LS}\right]$ is the life satisfaction score that is attached to death. The normalization by LS_d accounts for the fact that there exist states of life that are

considered not worth living (e.g., Ditto et al. 1996, Rubin et al. 2016). In this framework, an increase in longevity corresponds to a shift of A to A', where $A' > A$. This increase in longevity is considered to be welfare improving if $TLS' > TLS$, i.e. if overall life satisfaction of the representative agent increases.

One major drawback of this framework is its reliance on a measure of overall life satisfaction that rests on a very strong assumption: cardinality. This assumption is unlikely to hold for life satisfaction, which is typically measured on an ordinal scale in surveys. Therefore, I depart from this measure of overall life satisfaction and use a combination of two approaches to get as close as possible to this notion of overall life satisfaction without relying on the cardinality assumption. Initially, I use a novel approach. The time-to-death approach investigates changing end-of-life life satisfaction patterns over time. Although resting on a minimum set of assumptions, this approach can be sufficient to conclude that past increases in lifetime were welfare improving. In some cases, however, a complementary approach is required to weight increases in (satisfied) lifetime against losses in end-of-life life satisfaction. The life-expectancy approach suits this purpose, but requires some additional assumptions. Most importantly, it requires the choice of a cutoff value to distinguish between states of satisfaction and dissatisfaction. As the sensitivity of results to alternative cutoff values can be tested, however, this assumption is less restrictive than the cardinality assumption.

To understand the underlying idea of the time-to-death approach, it is useful to consider three important facts about the evolution of life satisfaction in old age:

1. *Stability-despite-loss-paradoxon:* Life satisfaction is relatively stable in old age despite aging-related losses (Diener et al. 1999, Kunzmann et al. 2000, Schilling 2006).

2. *Terminal decline:* Life satisfaction strongly declines in the last years of life. This decline is linear, possibly with a more pronounced drop in the last year of life (e.g., Gerstorf et al. 2008a,b, Palgi et al. 2010, Berg et al. 2011).

3. *Onset of terminal decline:* Life satisfaction starts to decline roughly three to five years before death (Gerstorf et al. 2008a,b, 2010).

Although it is yet unknown whether these facts also hold in aging societies across time, previous literature clearly suggests that it is crucial to investigate changes in end-of-life

life satisfaction over time to understand whether overall life satisfaction has increased with improved longevity in aging societies.

In principle, characteristics of the terminal life satisfaction decline might have changed with improved longevity. It is, for example, possible that the terminal decline extended, leading to much lower life satisfaction scores immediately before death. Alternatively, there might have been a shift in the onset of terminal decline or a change in the slope of terminal decline. Figure 2.2 illustrates some of these possible changes for the representative agent between 1985 and 2010 (figures on the left), and demonstrates how these changing end-of-life life satisfaction patterns over time can be uncovered by focusing on the last five years of life and analyzing time trends of average life satisfaction by time to death (figures on the right).[6] If I observe, for example, a downward sloping time trend of average life satisfaction for the last year of life, this is consistent with an extension of the terminal decline (see panel A, D). A change in the slope of terminal decline over time is reflected by narrowing (see panel C) or widening (see panel D) gaps between the average life satisfaction trends over time. Moreover, a shift in the onset of terminal decline can be deduced from any combination between the two. It is directly observable if the time trend of average life satisfaction for elderly who are four or five years prior to death is upward sloping or flat (see panel B).

A clear indication for a welfare improvement in terms of overall life satisfaction is the pattern of average life satisfaction trends that are upward sloping or flat over time. This is easily seen for the limiting case, which is depicted in panel B of Figure 2.2.[7] If the shift in the onset of terminal decline exactly corresponds to the shift in the age at death, the additional lifetime is exclusively spent in satisfaction. Further, considering that life satisfaction scores in the terminal decline phase are identical, it must be that overall life satisfaction of the representative agent increases. This holds even in the absence of the

[6] As usual, I compute averages to get from individuals to the representative agent of a society, i.e. I report time trends of average life satisfaction. Interestingly, time trends of the proportion of elderly in each state of satisfaction would yield the same conclusion, however. This is because in my data time trends of the proportion of elderly in states of satisfaction are decreasing or flat whereas time trends of the proportion of elderly in states of dissatisfaction are increasing or flat. Given that each time trend is consistent with a deterioration of life satisfaction over time, the overall result can be summarized in a single time trend of average life satisfaction.

[7] Note that panel B depicts the case that would result if the mentioned end-of-life life satisfaction facts hold in aging societies across time.

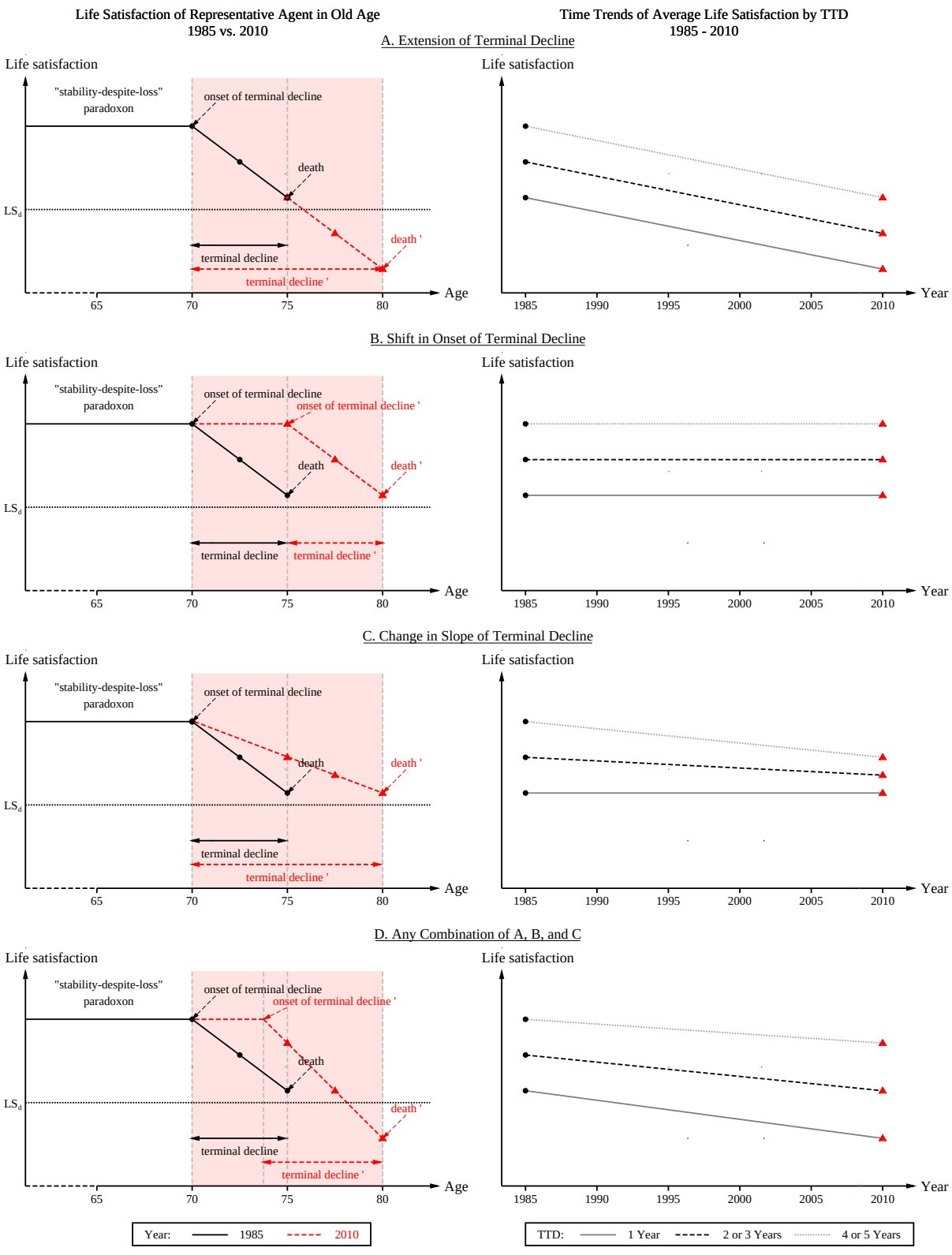

FIGURE 2.2. GRAPHICAL ILLUSTRATION OF THE TIME-TO-DEATH APPROACH

Notes: TTD = time to death. This figure illustrates how possible changes of life satisfaction of the representative agent in the last five years of life between 1985 and 2010 (figures on the left) translate to changing end-of-life life satisfaction patterns over time that can be uncovered by analyzing time trends of average life satisfaction by time to death (figures on the right). Own representation.

cardinality assumption.[8] In contrast, if average life satisfaction trends are downward sloping, this is a necessary but not sufficient condition for a welfare loss in terms of overall life satisfaction. The reason is that downward sloping time trends indicate a deterioration of the final period of life that may be compensated by an increase in satisfied lifetime. That is, overall life satisfaction may increase or decrease, depending on how losses in end-of-life life satisfaction are valued relative to gains in satisfied lifetime. Therefore, in this case a summary measure is required to obtain clear predictions.

I use satisfied life expectancy at age 60 as a summary measure that avoids the cardinality assumption. This summary measure indicates the expected number of satisfied life years beyond the age of 60. To weight increases in satisfied lifetime against increases in dissatisfied lifetime, I compare the proportion of expected satisfied lifetime to expected total lifetime at the age of 60 across time. That is, I make the following assumption to allow for welfare comparisons across time:

Assumption. *A satisfied life is a life where beyond the age of 60*

(i) the expected number of satisfied life years is high and

(ii) the proportion of the expected number of satisfied life years to the expected number of total life years is high,

and (ii) is of first-order importance.

Under this assumption, successful aging requires the proportion of expected satisfied lifetime to expected total lifetime beyond the age of 60 to be non-decreasing over time. People are willing to accept an extension of the dissatisfied lifetime at the end of their life provided that it is not too long. Moreover, they are willing to accept lower life satisfaction scores immediately before death provided that they are not too low. These are two important features that also find empirical support in discrete choice experiments on the willingness-to-pay for QALY gains stemming from life extensions (Pennington et al. 2015, Ahlert et al. 2016, Fischer et al. 2017).

[8] In the absence of the cardinality assumption, the magnitude of the welfare gain cannot be quantified, however. Under the cardinality assumption (and in continuous time), in panel B overall life satisfaction would correspond to the area between the age-satisfaction-curve and the horizontal line at LS_d, respectively. The welfare gain would correspond to the difference between the two areas for 2010 and 1985, i.e. the area between the black and the red satisfaction-age-curves, and the horizontal line at LS_d.

2.3 Data

This study uses data of the German Socio-Economic Panel. The SOEP is a nationally representative longitudinal study of households in Germany. It was launched in 1984. Initially, it included West German households only. After the German reunification, a representative sample of East German households was added. Currently, more than 20,000 adult residents are interviewed on an annual basis. The survey content includes rich information on demographics, employment, household composition, health, attitudes, and values. Information on overall life satisfaction is collected since 1984.[9]

The SOEP represents in many ways the most appropriate data set for this study. First, it provides information on overall life satisfaction for an extensive period of time, which coincides with substantial increases in life expectancy (see Figure 2.1). Second, the SOEP is very successful in following up its survey population.[10] Therefore, the overall sample is representative of the population living in private households (Kroh et al. 2008) and long-term care homes in Germany (Klein 1996). Third, providing reliable information on deaths, the SOEP is increasingly used for mortality-related analyses (e.g., Burkhauser et al. 2005, Gerstorf et al. 2010, Vogel et al. 2017). Fourth, the SOEP provides information on an extensive set of background characteristics, which can be used to rule out compositional changes in the population as an explanation for the observed changes in life satisfaction over time. One limitation of the SOEP, however, might be that the least healthy and the least satisfied are more likely to drop out, resulting in a selected sample of the elderly. Yet, Kroh (2014) shows that the primary reason for study dropout among older survey participants in the SOEP is mortality. Nevertheless, I address the problem of attrition in a sensitivity check later.

In the SOEP, information on overall life satisfaction is collected using the question "How satisfied are you with your life currently, all things considered?".[11] The answer to this question is measured on an eleven point Likert scale, ranging from 0 (very dissatisfied) to 10 (very satisfied). Life satisfaction is an evaluative measure of well-being, which is widely used in the economic literature (e.g., Di Tella et al. 2001, Frijters et al. 2004,

[9] For further information on the survey content and the sampling structure of the SOEP, see Wagner et al. (2007).

[10] Initial response rates were between 60% and 70%. Longitudinal attrition was also relatively low (about 15% for the second wave and less than 5% for subsequent waves). In order to explore the reasons for panel dropout, dropout studies were conducted in 1992, 2001, 2007, and 2008.

[11] In German: "Wie zufrieden sind Sie gegenwärtig, alles in allem, mit Ihrem Leben?"

Deaton 2008, Schwandt 2016). In contrast to hedonic (e.g., feeling of happiness, sadness, stress) or eudemonic measures of well-being (e.g., sense of meaning and purpose in life), it gives an assessment of the peoples' quality and goodness of lives (cf. Steptoe et al. 2015). Hence, it is particularly suitable for the purpose of this study.

To investigate how end-of-life life satisfaction changed over time, I use information about the mortality status and the year of death. This information was obtained either directly at the yearly interviews from remaining household members, relatives, and neighbors, or indirectly from official registries, which were contacted throughout dropout studies. Time to death is calculated by subtracting the survey year from the year of death.[12] Despite the strong attempts to follow-up study participants, deaths continue to be slightly underrepresented in the SOEP (Schnell and Trappmann 2006). Therefore, and because men are overrepresented among older SOEP participants, the SOEP slightly underestimates the average age at death, when compared to the official German death statistics (GBE 2016). Time trends of the average age at death are replicated quite well, however. To account for the observed level shifts, I use data of the official German death statistics and construct scale weights to adjust the estimates throughout the time-to-death approach.[13] To compute satisfied life expectancy at age 60, I use gender-specific life tables for West Germany in addition to the SOEP data. These period life tables are provided by the German Statistical Office on an annual basis since the late 1950s (German Statistical Office 2012a,b).

The main samples are comprised as follows: I focus on former West Germany because East Germany experienced several institutional and ideological changes after the German reunification, which renders East German data less appropriate for longitudinal compar-

[12]Lacking information on the month of death, it is impossible to determine the time to death more exactly. Consequently, a person who died in December 1980 and was surveyed in January 1978 is coded as being two years before death. The same holds true for a person who died in January 1980 and was surveyed in December 1978, although this person was much fewer months (13 months) prior to death at the time of the survey. On the contrary, a person who died in July 1980 and was surveyed in June 1979 is coded as being one year before death, although at the time of the survey this person was also 13 months prior to death. For reasons of simplicity, however, in what follows I refer to people in my sample as people who are one year, two or three years, and four or five years prior to death.

[13]To construct scale weights, subgroups in a given year were formed based on gender, five-year age-at-death intervals, and time to death. The oldest age-at-death interval represented an exception, encompassing people who died at age 90 or older. Population counts for deceased West German residents with German nationality were merged with sample counts for West German SOEP respondents by year of death and group characteristics, implicitly assuming that the composition of the population was constant in the five years before death.

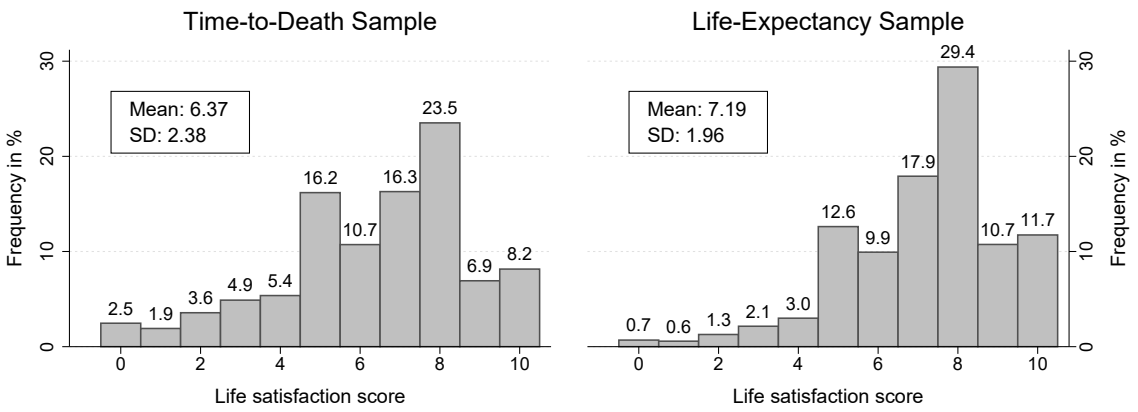

FIGURE 2.3. DISTRIBUTION OF LIFE SATISFACTION SCORES BY SAMPLE

Notes: Life satisfaction is measured on an eleven point Likert scale, ranging from 0 (very dissatisfied) to 10 (very satisfied). N = 9,371 (time-to-death sample) and N = 26,870 (life-expectancy sample).
Sources: SOEPv30 (1984-2013), own calculations.

isons.[14] I exclude migrants because the composition and the share of migrants changed over time. Finally, I restrict the analyses to the elderly, i.e. I focus on either respondents who died at age 60 or older and were within five years of death (time-to-death approach) or respondents who were aged 60 plus at the time of the survey interview irrespective of their remaining lifetime (life-expectancy approach). I use this age restriction to account for the fact that in the last three decades age-related deaths in West Germany mainly occurred at the age of 60 and at older ages.[15]

Figure 2.3 shows the distribution of life satisfaction scores for the time-to-death and life-expectancy samples. It illustrates for both samples that the life satisfaction distribution is highly left-skewed. For the life-expectancy sample, almost 50% of the responses are concentrated on the categories 7 and 8. In contrast, only 7.7% of the respondents report a life satisfaction score below 5 (the midpoint of the scale). The distribution of the time-to-death sample is less skewed than the distribution of the life-expectancy sample. The average satisfaction score lies with 6.37 more than three quarters of a Likert point below the average satisfaction score of the life-expectancy sample. The stronger

[14]Note that I only use data of the SOEP samples A to F. The samples A and B represent the original samples of West German households. The samples C to F were added at later stages to include East German households and to compensate for attrition.

[15]In this study, I am interested in age-related death as opposed to accidental death. According to the official German death statistics (GBE 2016), in West Germany less than 8% of the people with German nationality died before age 60 in 1985. In 2010, this share was even lower at 4.6%. As shifts in the legal retirement age across time may confound the results, I tested the robustness of results to a change in the age cutoff. I find that results are robust, if I use the age 65 as cutoff (results available upon request).

concentration of responses at lower life satisfaction scores for the time-to-death sample is consistent with a terminal decline in life satisfaction.

2.4 Time-to-Death Approach

In this section, I investigate how end-of-life life satisfaction changed with improved longevity over time. After providing some additional details on the estimation, I present the results of the time-to-death approach and show that they are robust to a large set of alternative explanations.

2.4.1 Empirical Strategy

I estimate time trends of average life satisfaction for 2,446 West German respondents who were within five years of death between 1985 and 2011. To uncover changes in both slope and onset of terminal life satisfaction decline, I distinguish between respondents who were one year, two or three years, and four or five years prior to death, and estimate time trends separately by time to death. To ensure a sufficiently large sample size and to smooth time trends slightly, I pool data of three years including and surrounding a survey year, when computing the averages for a survey year. Thus, estimates in a given year rely on roughly 200 to 400 observations (see Figure 2A.1 in Appendix 2A). I report weighted estimates throughout.

A meaningful interpretation of these estimates relies on the assumption that changes in end-of-life life satisfaction over time can only be attributed to increased longevity. This is a very strong assumption, since there probably exist other time-varying factors that contribute to changes in end-of-life life satisfaction over time. Compositional effects such as, for example, increased income or education may lead to a more satisfied population over time because both income and education are (weakly) positively related to life satisfaction (e.g., Oswald 1997, Frey and Stutzer 2002, Clark et al. 2008). In contrast, cohort effects may lead to declining life satisfaction scores over time because of Germany's unique history during and after World War II, although empirical evidence in this regard is mixed (Gwozdz and Sousa-Poza 2010, Baird et al. 2010).[16] In addition, time-in-panel effects may lead to declining life satisfaction scores over time as reported life satisfaction is negatively

[16]Due to their war experience, earlier born cohorts may more positively assess their current life than later born cohorts when making intrapersonal comparisons.

related to the duration spent in a panel (e.g., Kassenboehmer and Haisken-DeNew 2012, Baetschmann 2014).[17] Therefore, I explore the role of alternative explanations using a broad set of sensitivity checks. I discuss the sensitivity checks, after presenting the main results of the time-to-death approach.

2.4.2 Results

Figure 2.4 depicts time trends of average life satisfaction and average age at death for West German elderly without migration background who were one year (solid line in dark gray), two or three years (dashed-dotted line in black), and four or five years (dashed line in light gray) prior to death.[18] This figure clearly suggests that the last five years of life, on average, deteriorated with improved longevity over time. The figure on the left indicates that average life satisfaction prior to death strongly declined over time, irrespective of the time to death. Between 1985 and 2010, average life satisfaction decreased by almost one Likert point. This decline in life satisfaction reaches statistical significance, and it is large, when compared to the change in life satisfaction that is caused by a change in alternative respondents' characteristics such as education or the employment status (e.g., Oreopoulos (2007) for education and Clark and Oswald (1994), Winkelmann and Winkelmann (1995), and Kassenboehmer and Haisken-DeNew (2009) for unemployment and job loss). The figure on the right demonstrates that the decline in life satisfaction went along with an increase in longevity.[19]

Figure 2.4 also suggests additional deteriorations of life satisfaction in old age that go beyond the last five years of life. The figure on the left shows that the terminal life satisfaction decline also holds in aging societies across time. In each given year, life satisfaction declines, on average, with proximity to death. Yet, improvements in

[17]The former authors suggest that such a finding is consistent with increasing confidence in the interviewer and, hence, more honest (in this case lower) life satisfaction answers over time.

[18]Note that each trend line stops at a different point in time as a SOEP respondent's death is only observed until the mid of 2013. Respondents who died in 2013 were one year prior to death in 2012, three years prior to death in 2010, and five years prior to death in 2008. To avoid compositional changes in the study population across years when computing three-year averages, the corresponding trend lines end in 2011, 2009, and 2007.

[19]In a balanced panel (i.e., in the absence of attrition and sample refreshment), time trends of the average age at death should be identical in shape. They would simply be shifted to the right with increasing proximity to death because people who were, for example, four or five years prior to death in 1985 correspond to those who were two or three years prior to death in 1987. The fact that I find time trends of the average age at death that are parallel to each other (also pre-weighting) suggests for the SOEP that attrition with respect to age is neither increasing nor decreasing in the final years of life over time, despite improved longevity.

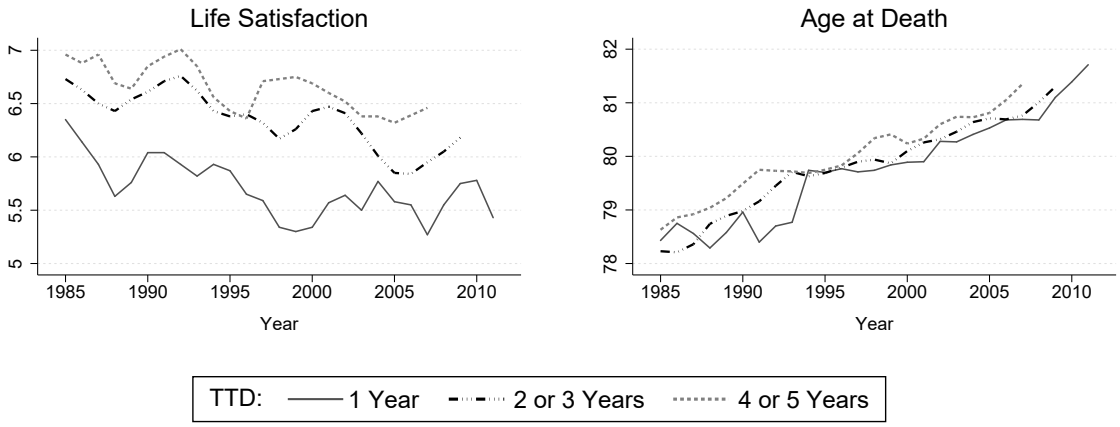

Figure 2.4. Time Trends of Average Life Satisfaction and Average Age at Death by Time to Death

Notes: TTD = time to death. Three-year averages are estimated based on 26,264 respondent-year-observations, which are obtained from 2,446 deceased respondents. Life satisfaction is measured on an eleven point Likert scale, ranging from 0 (very dissatisfied) to 10 (very satisfied). Age at death is measured in years.
Sources: SOEPv30 (1984-2013) and GBE (2016), own calculations.

longevity came with changing characteristics of the terminal life satisfaction decline that are consistent with an extension of the terminal decline phase. The slope of terminal life satisfaction decline became somewhat smaller over time because average life satisfaction continuously decreased for West German elderly who were more than one year prior to death, whereas it only decreased throughout the first decade and stayed relatively constant thereafter for West German elderly who were one year prior to death. Moreover, average life satisfaction four or five years prior to death continuously decreased over time. Both suggests that a possible shift in the onset of terminal decline was smaller than the shift in the age at death. Hence, it is very likely that the dissatisfied period at the end of people's life, on average, extended with improved longevity for West German elderly.

2.4.3 Sensitivity Tests

In this section, I test the sensitivity of the time-to-death results to alternative explanations. I show that compositional effects, cohort effects, time-in-panel effects, a simple aging-unrelated time trend in life satisfaction, endogenous onset of disease and terminal life satisfaction decline, an increasingly negatively selected sample due to the age restriction, and attrition cannot explain the decline in end-of-life life satisfaction over time. Moreover, I show that results are robust to the weighting strategy. I discuss each sensi-

tivity check in more detail below. The results of the sensitivity checks are presented in Appendix 2A.

To assess the role of compositional effects, cohort effects, and time-in-panel effects, I estimated weighted individual-level life satisfaction regressions by time to death. Using the pooled three-year average data set, I regressed life satisfaction on a set of year dummies, subsequently adding distinct sets of controls to the regressions: To account for compositional effects, I added a dummy for males, years of education, net household income, dummies for the interview month, and dummies for the state of residence; to account for cohort effects, I added five-year cohort dummies; and to account for time-in-panel effects, I added a linear term for time-in-panel duration.[20] Figure 2A.2 graphically depicts the year dummy coefficient estimates of these regressions, which are relative to the year 1985. This figure demonstrates that the decline in end-of-life life satisfaction persists, even after controlling for compositional effects, cohort effects, and time-in-panel effects, although there is some evidence that time-in-panel effects contributed to the decline in end-of-life life satisfaction over time.[21]

To investigate whether the decline in end-of-life life satisfaction over time represents a general aging-unrelated time trend, I used a difference-in-difference type of approach. More specifically, I estimated time trends of average life satisfaction for two distinct control groups of young West German respondents. Life satisfaction of young respondents should not be affected by a shift in age-related death because neither they nor their parents are close to age-related death. Under the assumption that in the absence of a shift in age-related death young and old West Germans (the latter being close to death) exhibit the same life satisfaction trends, the difference between the trend lines of these two groups gives the impact of increased longevity on end-of-life life satisfaction. Hence, in order to rule out that a general time trend contributed to the decline in end-of-life life satisfaction over time, life satisfaction trends for young West Germans should be upward sloping or

[20]Note that the classical identification problem between age, year, and cohort does not arise as age is not included in these regressions. The coefficient on time-in-panel duration is identified due to sample refreshment and non-response. Household income is deflated and need-weighted, i.e. it adjusts for purchasing power, and household size using modified OECD equivalence weights.

[21]This result also holds if I control for relative as opposed to absolute income (e.g., poverty indicators, distance to median income). Moreover, I find that an urban-rural drift in the place of living is unlikely to account for the decline in end-of-life life satisfaction. Information on community size is, however, only available in SOEP since 1995 and, thus, not controlled for in the regressions presented here.

flat.[22] Figure 2A.3 shows that there is at most a small decline in average life satisfaction of young West Germans. This decline vanishes upon controlling for time-in-panel effects, suggesting that a general time trend does not explain the decline in end-of-life life satisfaction.[23]

Previous literature suggests that more satisfied people tend to live longer (e.g., Veenhoven (2008) for a review). Reverse causality is problematic for two reasons. First, the age restriction for the time-to-death sample may lead to a decreasingly positively selected sample over time. Second, at the individual level onset of disease, onset of terminal life satisfaction decline, and age at death are endogenous. To address the concern of decreasing sample selectivity, I dropped the age restriction and reestimated average life satisfaction trends by time to death for all West German adults without migration backgrounds who were within five years of death between 1985 and 2011. Figure 2A.4 shows that average life satisfaction trends exhibit the same slopes of decline, thereby ruling out this potential explanation. To address the concern of endogeneity at the individual level, I reestimated average life satisfaction trends using objective death probabilities instead of actual distance to death.[24] Objective death probabilities were retrieved from period life tables for West Germany (German Statistical Office 2012a,b) and merged by gender, age, and year with the SOEP data. Within each death probability group the fraction of people who were within their last years of life is constant over time, but due to the rectangularization of survival curves the average age increases within each death probability group over time. Hence, average life satisfaction trends should by downward sloping between 1985 and 2010 if population aging rather than endogenous shifts in the onset of terminal life satisfaction decline contribute to the decline in end-of-life life satisfaction over time. Figure 2A.5 shows for West Germans aged 60 and older that this is indeed the case.[25]

To investigate whether changed attrition patterns over time contributed to the downward sloping end-of-life life satisfaction trends, I estimated a linear probability model for study dropout using the unbalanced time-to-death sample. Table 2A.2 reports the regres-

[22] A general time trend results from (unobserved) time-varying factors that affect life satisfaction of treatment and control groups in the same fashion (e.g., increase in GDP, technological progress, change in nutritional habits, improvements in access to health care).

[23] Using this strategy, it is impossible to rule out age-dependent time trends, however.

[24] Unlike actual distance to death, objective death probabilities are unrelated with life satisfaction and onset of disease at the individual level.

[25] Note that consistent with the terminal life satisfaction decline, in each year life satisfaction reduces, on average, with objective death probability.

sion results. It shows for the last five years of life that less satisfied West German elderly are more likely to drop out of the SOEP, but that the attrition pattern with respect to life satisfaction did not change over time, as indicated by the coefficient on the interaction term between the post-1995 dummy and life satisfaction.

Finally, Figure 2A.6 illustrates the effect of weighting for the main variable of interest and the two variables that were used to adjust the SOEP data to the official German death statistics. This figure clearly shows that the time-to-death results are not affected by the use of scale weights, although post-weighting the average age at death is slightly higher and the share of men is more than five percentage points lower in any given year.

2.5 Life-Expectancy Approach

In this section, I investigate whether increases in satisfied lifetime compensated for the decline in end-of-life life satisfaction over time such that people became better off in terms of overall life satisfaction with improved longevity. After providing details on the estimation strategy, I present the results of the life-expectancy approach. Then, I again show that the results are robust to alternative explanations.

2.5.1 Empirical Strategy

I estimate time trends of satisfied life expectancy at age 60. Satisfied life expectancy at age 60 is estimated based on Sullivan's method (Sullivan 1971). This method is widely used in the sociological literature to compute healthy life expectancies (Jagger and Robine 2011). It has three major advantages over multistate life table methods[26]: First, it is easier to implement. Second, it has lower data requirements. Third, it relies on fewer assumptions (cf. Imai and Soneji 2007). Contrary to multistate life table methods, however, Sullivan's method (if based on period life tables) relies on the assumption that age-specific mortality and satisfaction prevalence rates are constant over time. This assumption is unlikely to hold for West Germany where mortality and satisfaction-prevalence rates at all ages have continuously changed since 1985. Yet, Mathers and Robine (1997) show in a simulation study that both methods yield similar estimates if changes over time are smooth and occur regularly. As illustrated by Figures 2B.1 and 2B.2 in Appendix 2B,

[26]Multistate life table methods model transitions into different states. That is, they use flow rather than stock data to estimate satisfied life expectancy.

there is no evidence for sudden changes in age-specific mortality and satisfaction prevalence rates for West Germany over time.[27]

The idea of Sullivan's method is to divide total life expectancy at age 60 into satisfied and dissatisfied life expectancy at age 60 by combining data from two different sources. While the person-years lived in each age interval are obtained from period life tables, satisfaction prevalence rates for the corresponding age intervals are estimated based on survey data and then used to weight the person-years lived in each age interval. After weighting, the computation of satisfied life expectancy is equivalent to the one of standard life expectancy, i.e. it is computed by summing over all the weighted person-years lived from age 60 onwards and dividing the weighted total person-years lived at age 60 by the number of 60-year-old survivors. Hence, Sullivan's estimator for satisfied life expectancy at age 60 is formally defined as

$$\hat{e}^s_{60} = \frac{\sum_{x \in A_{60}} \hat{h}^s_{x,n_x} L_{x,n_x}}{l_{60}}, \tag{2.2}$$

where A_{60} represents the set of starting ages x such that $x \geq 60$, $\hat{h}^s_{x,n_x}$ denotes the sample fraction of satisfied survey respondents in the age interval $[x, x + n_x)$, L_{x,n_x} indicates the person-years lived in the age interval $[x, x + n_x)$, and l_{60} gives the number of 60-year-old survivors. Dissatisfied life expectancy at age 60 is estimated either by replacing $\hat{h}^s_{x,n_x}$ with the sample fraction of dissatisfied survey respondents, $\hat{h}^{ds}_{x,n_x} = 1 - \hat{h}^s_{x,n_x}$, in equation (2.2) or by directly subtracting satisfied life expectancy at age 60, $\hat{e}^s_{60}$, from total life expectancy at age 60, $\hat{e}_{60}$.

I use five-year age intervals, i.e. $n_x = n = 5$ for all but the last age interval, which ranges from 85 to the oldest observed age. Therefore, the single-year unabridged life tables are transformed into five-year abridged life tables.[28] Gender-age-specific satisfaction prevalence rates are estimated by the gender-age-specific sample fractions of SOEP respondents in any of the three states: dissatisfied (life satisfaction of 0 to 6), moderately

[27] In particular, the mortality rates changed very smoothly. Between 1985 and 2010, mortality decreased continuously for all age groups. Changes in satisfaction prevalence rates were also relatively smooth, although estimates are more volatile due to the smaller sample size.

[28] For the sake of consistency, I again focus on West German residents without migration backgrounds. As period life tables for West Germany are calculated based on residents with and without migration backgrounds, I implicitly assume the same age-specific mortality rates across immigrants and non-immigrants. In addition, I implicitly assume equivalent age-specific mortality rates for West Berlin and the other West German states after 1999, since West Berlin was excluded from the calculations for West Germany since 2000.

satisfied (life satisfaction of 7 or 8), and very satisfied (life satisfaction of 9 or 10). The threshold values correspond to the first and third quartile of life satisfaction in the life-expectancy sample. I show later, however, that the results are robust if I use the midpoint of the Likert scale to distinguish between states of satisfaction and dissatisfaction. Consistent with the data obtained from period life tables, I pool data of three years including and surrounding the survey year when estimating the satisfaction prevalence rates for a given year.[29] Moreover, due to the sampling structure of the SOEP, I use cross-sectional survey weights in the estimations.

To allow for comparisons of satisfied life expectancy at age 60 over time, I report estimates for the years 1985, 1990, 2000, and 2010 in both absolute (number of years) and relative (proportion of life) terms. To test for significant differences over time, I derive standard errors based on the delta method (see Appendix 2C). These standard errors differ from those that are usually derived in the literature for repeated cross-sectional data. I account for serial correlation across observations of the same respondent because I use longitudinal data and pool the data of three years when estimating the satisfaction prevalence rates for a given year. For testing purposes, I have to assume, however, that the covariance between two satisfied life expectancy estimates can be ignored as I independently computed the standard errors of satisfied life expectancy across time. This assumption is most plausible for the largest time difference: Less than five percent of the observations between 2009 and 2011 are from elderly respondents who also took part in the SOEP surveys between 1984 and 1986. Therefore, I only report test results for the difference of satisfied life expectancy between 2010 and 1985.

2.5.2 Results

Table 2.1 presents the estimates of very satisfied, moderately satisfied, and dissatisfied life expectancy at age 60 for West Germans without migration backgrounds by gender and year. Beyond the absolute values in years (rows 2 to 4), this table shows the proportions on total life expectancy in percent (rows 5 to 7). The first row of Table 2.1 indicates the increase in total life expectancy at age 60 for both men and women, which was depicted in Figure 2.1. Between 1985 and 2010, life expectancy for a 60 year-old man increased by

[29] The sample sizes used to estimate the gender-age-specific satisfaction prevalence rates in a given year range from 120 to 1,546 observations. Two exceptions are the sample sizes for men aged 85 and older in 1985 and 1990. With 61 and 85 observations, these sample sizes are much smaller.

Table 2.1. Total and Satisfied Life Expectancy at Age 60 by Gender and Year, West Germany (1985 – 2010)

	Men					Women				
	1985	1990	2000	2010	Difference (2010-1985)	1985	1990	2000	2010	Difference (2010-1985)
Total LE	17.1	17.8	19.7	21.4		21.6	22.2	23.9	25.0	
Very satisfied LE	5.5	4.9	4.1	3.5	-2.0***	6.9	5.4	4.9	4.2	-2.7***
	(0.45)	(0.43)	(0.31)	(0.42)	(0.62)	(0.39)	(0.42)	(0.29)	(0.33)	(0.51)
Moderately satisfied LE	6.9	7.9	10.2	11.3	4.4***	8.0	9.4	10.9	12.2	4.2***
	(0.43)	(0.41)	(0.42)	(0.45)	(0.62)	(0.36)	(0.41)	(0.35)	(0.39)	(0.53)
Dissatisfied LE	4.7	5.0	5.4	6.6	1.9***	6.6	7.4	8.1	8.7	2.0***
	(0.43)	(0.45)	(0.42)	(0.45)	(0.62)	(0.38)	(0.44)	(0.38)	(0.41)	(0.56)
Proportion of ... to total LE										
Very satisfied LE	32.2	27.7	20.6	16.3	-15.9***	32.1	24.2	20.5	16.8	-15.3***
	(2.66)	(2.39)	(1.58)	(1.95)	(3.30)	(1.82)	(1.91)	(1.23)	(1.31)	(2.24)
Moderately satisfied LE	40.1	44.4	52.1	52.7	12.6***	37.1	42.3	45.7	48.6	11.5***
	(2.49)	(2.31)	(2.12)	(2.12)	(3.27)	(1.65)	(1.86)	(1.46)	(1.57)	(2.28)
Dissatisfied LE	27.7	28.0	27.3	30.9	3.2	30.8	33.5	33.8	34.6	3.80**
	(2.49)	(2.55)	(2.13)	(2.09)	(3.25)	(1.76)	(2.00)	(1.60)	(1.65)	(2.41)
N	2,071	1,867	4,107	3,807		3,084	2,638	5,069	4,227	

Notes: LE = life expectancy. State-dependent life expectancies at age 60 are calculated based on Sullivan's method. They indicate the number of years that a 60-year-old survivor can expect to live in any of the three states: very satisfied (life satisfaction of 9 or 10), moderately satisfied (life satisfaction of 7 or 8), and dissatisfied (life satisfaction of 0 to 6). The column "Difference" indicates the change in state-dependent life expectancy between 1985 and 2010 and shows whether this change is statistically different from zero. The change is measured either in years (absolute values) or in percentage points (proportions). Due to rounding errors a subtraction of the 2010 value from the 1985 value is not always equal to the indicated difference. Moreover, proportions do not always add up to 100 percent. Standard errors that allow for serial correlation across observations of the same respondent are reported in parentheses.

** p<0.05 *** p<0.01.

Sources: SOEPv30 (1984-2011) and (German Statistical Office, 2012a,b), own calculations.

4.3 years, while life expectancy for a 60 year-old woman increased by 3.5 years. Despite the faster growth of life expectancy for men, in 2010 life expectancy at age 60 was still 3.6 years lower for men than women.

Very satisfied life expectancy at age 60 decreased over time, whereas both moderately satisfied life expectancy and dissatisfied life expectancy increased over time. Between 1985 and 2010, the corresponding changes are highly statistically significant. In 2010, only half of the additional lifetime was, on average, spent in states of satisfaction. This suggests that the onset of terminal decline shifted with improved longevity to older ages, but that this shift was smaller than the shift in the age at death. The increase in dissatisfied life expectancy by two years between 1985 and 2010 reflects both the extension of dissatisfied lifetime at the end of human life and the larger fraction of dissatisfied West German elderly five years before death. These life-expectancy results are in line with the time-to-death results, which also pointed towards an extension and deterioration of the terminal decline phase.[30]

The relatively strong increase in dissatisfied life expectancy relative to satisfied life expectancy at age 60 is reflected in an increased proportion of expected dissatisfied lifetime to expected total lifetime. While in 1985 a 60 year-old man was expected to live 27.7% of his remaining life in states of dissatisfaction, this proportion increased to 30.9% in 2010. For a 60 year-old woman the corresponding proportions were 30.8% in 1985 and 34.6% in 2010. The increase in the proportion of expected dissatisfied lifetime only reaches statistical significance for women, however. Yet, overall, the analysis of satisfied life expectancy at age 60 suggests that the increase in satisfied lifetime did not compensate for the extension and deterioration of the dissatisfied period at the end of West German elderly's lives. That is, the overall quality of life has, on average, deteriorated with improved longevity in West Germany since 1985.

To understand what contributed to the change of satisfied life expectancy at age 60 over time, I estimated counterfactual satisfied life expectancy at age 60, keeping age-specific mortality rates constant as of 1985. This allows me to distinguish between changes that result from direct declines in mortality and changes that result from both indirect

[30]Between 1985 and 2010, both the increase in dissatisfied life expectancy and the decrease in very satisfied life expectancy were more pronounced for women than men. These gender differences might be explained by a level effect: Between 1985 and 2010, women turned, on average, 3.6 to 4.5 years older than men conditional on surviving to age 60. As the quality of life is worse at very old ages, which were more likely to be reached by women than men, this can explain why women were less satisfied with the additional lifetime in 2010 than men.

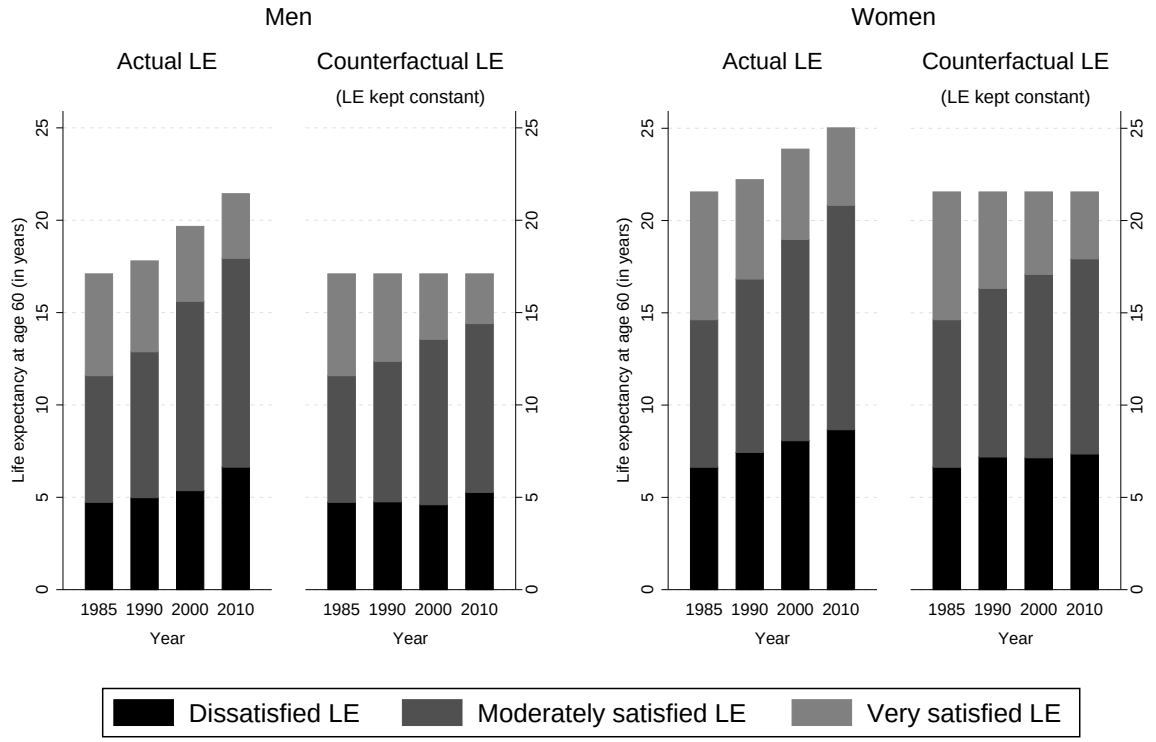

Figure 2.5. Actual and Counterfactual Satisfied Life Expectancy at Age 60 by Gender and Year

Notes: LE = life expectancy. These figures show the evolution of total and satisfied life expectancy at age 60 in West Germany by gender. Figures on the left each illustrate changes of actual life expectancy (resulting from changes in both satisfaction prevalence *and* mortality), whereas figures on the right each illustrate changes of counterfactual life expectancy (resulting mainly from changes in satisfaction prevalence). Total life expectancy is divided into the number of years that a 60-year-old survivor can expect to live in the very satisfied (life satisfaction of 9 or 10), moderately satisfied (life satisfaction of 7 or 8), and dissatisfied (life satisfaction of 0 to 6) states.
Sources: SOEPv30 (1984-2011) and German Statistical Office 2012a,b, own calculations.

declines in mortality *and* changes in satisfaction prevalence.[31] Figure 2.5 presents the estimates of counterfactual life expectancy at age 60 by gender and contrasts them with the estimates of actual life expectancy at age 60 that were presented in Table 2.1. At each point in time, total life expectancy – as indicated by the full length of a bar – is divided into the number of years that a 60 year-old survivor can expect to live in the very satisfied, moderately satisfied, and dissatisfied states. Changes in actual state-dependent

[31]If I only were to vary mortality rates, I clearly capture the effect of population aging because upon adjusting the number of survivors at all ages I would move along the age-satisfaction-prevalence curve. Yet, I would neglect the effect that declining mortality may have on satisfaction prevalence, possibly overestimating the negative effect of population aging. If, for example, with increased longevity the onset of terminal decline shifted for the average 60-year old by two years to age 75, life satisfaction among the average elderly aged 73 to 75 would increase as fewer elderly would be in their terminal decline phase. Such an indirect mortality effect can tilt the age-satisfaction curve in a given age range upwards and would be captured once varying the satisfaction prevalence rates over time. Therefore, if I only were to vary satisfaction prevalence rates, I would capture the effects of shifting *and* tilting age-satisfaction curves.

life expectancy, which result from changes in satisfaction prevalence and indirect declines in mortality, are shown within the left figure, respectively. In contrast, changes that result solely from direct declines in mortality are assessed by fixing a given year and comparing actual and counterfactual satisfied life expectancy across the left and right figures, respectively.

A comparison of the counterfactual estimates within the right figures suggests that the observed decrease of very satisfied life expectancy between 1985 and 2010 was largely attributable to a decline in satisfaction prevalence over time, as reflected by the strong decrease of counterfactual very satisfied life expectancy over time. The direct effect of declining mortality does not contribute to the decline of very satisfied life expectancy. It is clearly seen for any given year that very satisfied life expectancy in the left figures hardly changes, when compared to its counterfactuals in the right figures. The opposite holds true for the changes in dissatisfied life expectancy. Almost the full increase in dissatisfied life expectancy at age 60 between 1985 and 2010 is explained by direct declines in mortality. Changes in satisfaction prevalence over time contributed, if at all, only very little to the increase of dissatisfied life expectancy over time. The latter is crucial because it suggests that explanations other than improved longevity are unlikely to explain the increase in dissatisfied life expectancy.

2.5.3 Sensitivity Tests

In this section, I test the sensitivity of the life-expectancy results to alternative classification rules for the distinction between states of satisfaction and dissatisfaction, cohort effects, time-in-panel effects, and the use of cross-sectional survey weights. I show that results are robust to all but one modification. If I use an equal point split classification, I no longer find an increase in dissatisfied life expectancy between 1985 and 2010. Although this classification comes closest to the classification used in Yang (2008), this classification is problematic because it pools states of satisfaction and dissatisfaction in the intermediate category, thereby hiding important patterns over time. The results of the sensitivity checks are shown in Appendix 2B and briefly discussed below.

I investigated the robustness of results to two alternative classification schemes: First, I used the midpoint of the eleven point Likert scale to distinguish between dissatisfied and moderately satisfied states, while keeping the threshold value for the very satisfied

state constant. Second, I used an equal point split classification to distinguish between dissatisfied, moderately satisfied, and very satisfied states. Figure 2B.3 depicts the results. Under the midpoint split classification (top figures), I continue to find a strong increase of dissatisfied life expectancy at age 60 in both absolute and relative terms. This does not hold true, however, for the equal point split classification (bottom figures).[32] Under this classification, I essentially replicate the findings of Perenboom et al. (2004) and Yang (2008), i.e. I no longer find an increase of dissatisfied life expectancy over time. This study suggests that the primary reason is that the equal point split classification with three states of satisfaction pools states of satisfaction and dissatisfaction in the intermediate category, which makes it impossible to uncover the extension of dissatisfied lifetime.[33]

To investigate whether cohort effects and time-in-panel effects contributed to the change of satisfied life expectancy at age 60 over time, I reestimated actual and counterfactual life expectancies after adjusting satisfaction prevalence rates for cohort effects and time-in-panel effects. Figures 2B.4 and 2B.5 show that the life-expectancy results continue to hold, even after accounting for cohort effects or time-in-panel effects. Unlike cohort effects, time-in-panel effects contributed to the changes in very satisfied life expectancy, but neither of them contributed to the increase in dissatisfied life expectancy at age 60 over time, which is in line with the previous interpretation of results.

Finally, Figure 2B.6 demonstrates that weighted and unweighted results lead to identical conclusions.

2.6 Mechanisms

This section discusses two important mechanisms for the decline in end-of-life life satisfaction over time: health and social isolation. Both of them can explain why increases in lifetime came along with substantial losses of life satisfaction in old age.

2.6.1 Health

Health gradually declines with age (DePinho 2000, Rosenthal and Kavic 2004) and poor health is negatively related to life satisfaction (e.g., Oswald and Powdthavee 2008). Thus,

[32] Under this classification rule only 7.7% of the observations are classified as dissatisfied (cf. Figure 2.3).

[33] Note that the conclusion with respect to what contributes to the changes of satisfied life expectancy over time are robust to both variations of the classification rule (cf. Figure 2B.3, right figures).

one obvious channel through which increased longevity can affect life satisfaction is health. Whether or not it does primarily depends on medical technological progress. If medical technological progress shifted the onset of disease by more than the age at death, or, if it significantly improved the quality of life with disease (cf. Fries (1980): compression of morbidity hypothesis), health cannot account for the observed decline in life satisfaction over time. However, if medical technological progress mainly extended the lifetime with disease (cf. Gruenberg (1977) and Olshansky et al. (1991): expansion of morbidity hypothesis), poor health is very likely to be an important mechanism.

Figure 2.6 shows time trends of the main health indicators in the SOEP by time to death. It suggests that a deterioration of health likely is an important mechanism. Three out of the four objective health indicators show a sharp rise with increased longevity (middle and bottom figures). The share of elderly within five years of death who had a severe disability more than doubled between 1985 and 2010. In 2010, almost every second elderly exhibited a health impairment immediately before death. This increase in disability went along with a strong increase in the share of elderly who lived in residential homes. A very important reason for living in a residential home are health impairments. Hence, unsurprisingly, roughly one third of the elderly in residential homes reported receiving compulsory nursing care insurance that depends on the disability status. The average number of hospitalizations in the year preceding the interview also grew substantially (100% increase), but mainly for those who were one year or four to five years before death. This increase was predominantly driven by an increase in the probability of hospitalization. Solely the number of doctor visits within the three months preceding the interview showed an improvement over time, but this may be, for example, for cost-cutting reasons throughout the last decades. In contrast, subjective health measures (top figures) were more or less constant over time. One likely explanation for the stability of subjective health measures are interpersonal comparisons (Steffel and Oppenheimer 2009). If people assess their health relative to people in the same age group and health declines for all people in this reference group by the same amount, people will indicate the same health status as they did before the decline, despite their worsened health.[34]

[34]Unlike the overall quality of life, health is easily compared across people. This may explain why there is no decline in end-of-life health satisfaction over time, but a decline in end-of-life life satisfaction.

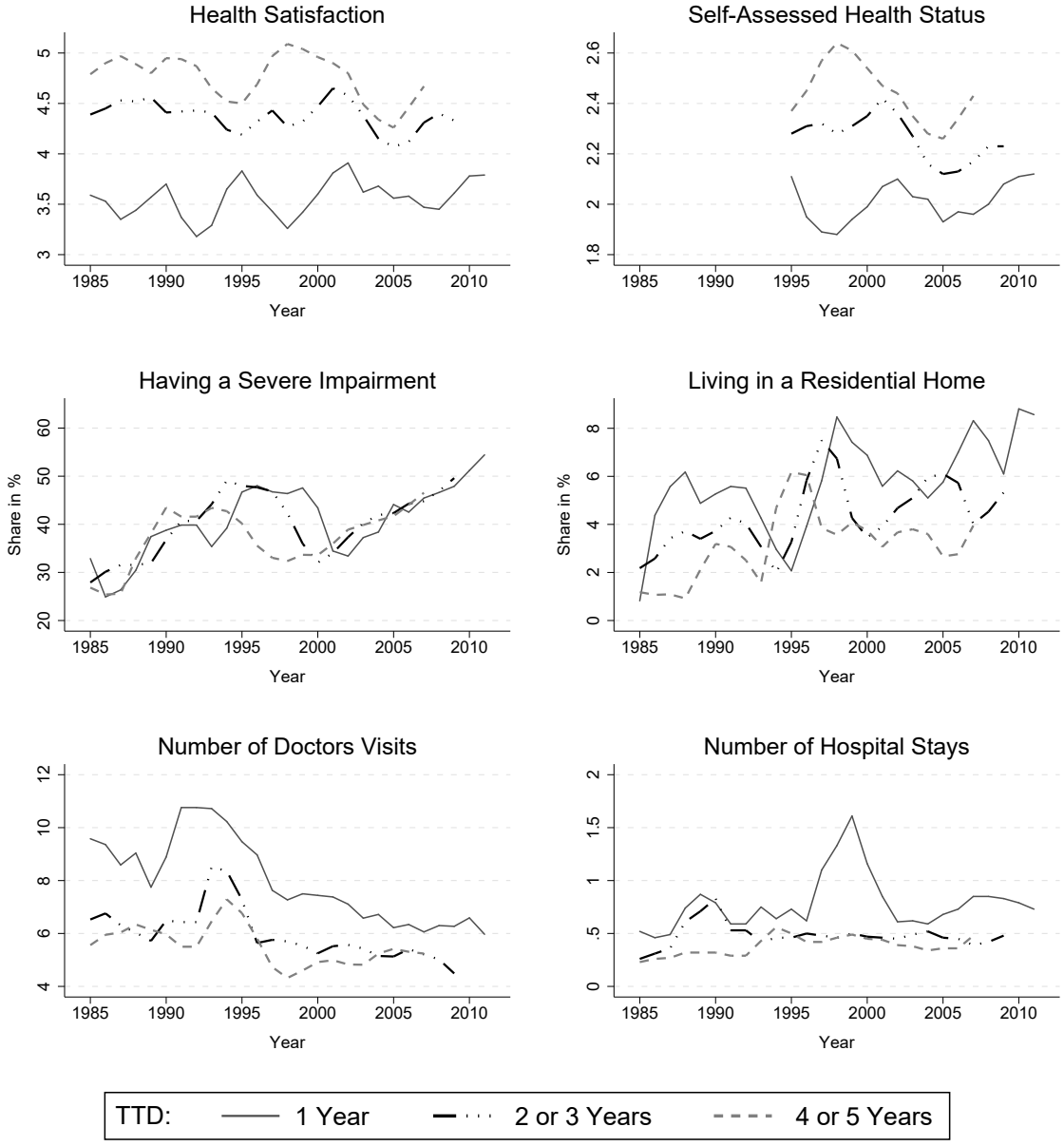

Figure 2.6. Time Trends of Average Health Indicators by Time to Death

Notes: TTD = time to death. Three-year averages are estimated based on 26,264 respondent-year-observations, which are obtained from 2,446 deceased respondents. Due to missing information (information not collected in 1990 and 1993) and item-non-response, three-year averages for the self-assessed health status, the share of elderly with severe impairments, the number of doctor visits in the last three months, and the number of hospital stays in the previous year are based on fewer respondent-year-observations. Health satisfaction is measured on an eleven point Likert scale, ranging from 0 (very dissatisfied) to 10 (very satisfied). The self-assessed health status is measured on a five point scale, ranging from 1 (bad) to 5 (very good). Information on the self-assessed health status is collected on an annual basis since 1994. Therefore, the first data point is depicted in 1995.
Sources: SOEPv30 (1984-2013) and GBE (2016), own calculations.

Figure 2D.1 in Appendix 2D provides further evidence for the health channel.[35] It demonstrates that the decline in end-of-life life satisfaction over time becomes much smaller after controlling for health indicators in individual-level life satisfaction. Health satisfaction explains roughly one third of the decline in life satisfaction of respondents who were two or three years before death, while the disability status fully explains the decline in life satisfaction of respondents who were four or five years before death. The adjusted R-squared more than doubles upon the inclusion of these health indicators. Solely for respondents who were one year before death the deterioration of health did not contribute to the decline in life satisfaction over time. Although this speaks against a health mechanism for the last year of life, this result is plausible. The last year of life was repeatedly shown to be associated with high medical spending (e.g., Zweifel et al. 1999, Seshamani and Gray 2004), which is indicative of very poor health. Thus, further health deteriorations are unlikely to matter for life satisfaction in the last year of life. Overall, these results support the expansion of morbidity hypothesis.

2.6.2 Social Isolation

Social isolation and inactivity are negatively associated with life satisfaction (e.g., Chappell and Badger 1989, Pinquart and Sörensen 2000, Powdthavee 2008), also prior to death (Gerstorf et al. 2016). Among various measures of social isolation, disconnectedness with social peers and a low number of friends have the most detrimental impact on life satisfaction (Chappell and Badger 1989, Pinquart and Sörensen 2000). Due to a reduction of multigenerational households (German Statistical Office 2016b) and increased geographical distance between adult children and elderly parents in Germany (e.g., Mahne and Huxhold 2017), fewer personal contacts with family members likely contributed to the decline in end-of-life life satisfaction over time. Moreover, given the increased variation in longevity at age 60 in industrialized countries (Engelman et al. 2010), it is possible that West German elderly became more likely to experience a friend's death early in life, reducing the frequency of personal contacts with friends in the final period of life over time.[36] In addition, fewer personal contacts with family and friends might have resulted

[35] Table 2D.1 in Appendix 2D shows the corresponding regression results.

[36] Figure 2D.2 in Appendix 2D replicates the finding of Engelman et al. (2010) for West Germany. The figure on the left shows that variation in longevity at age 60 strongly increased for the life-expectancy sample until the early 2000s. Thereafter, it started to decline, however. The figure on the right uses data of the official German death statistics (German Statistical Office 2016a) to demonstrate that the same pattern is found for all Germans in official data.

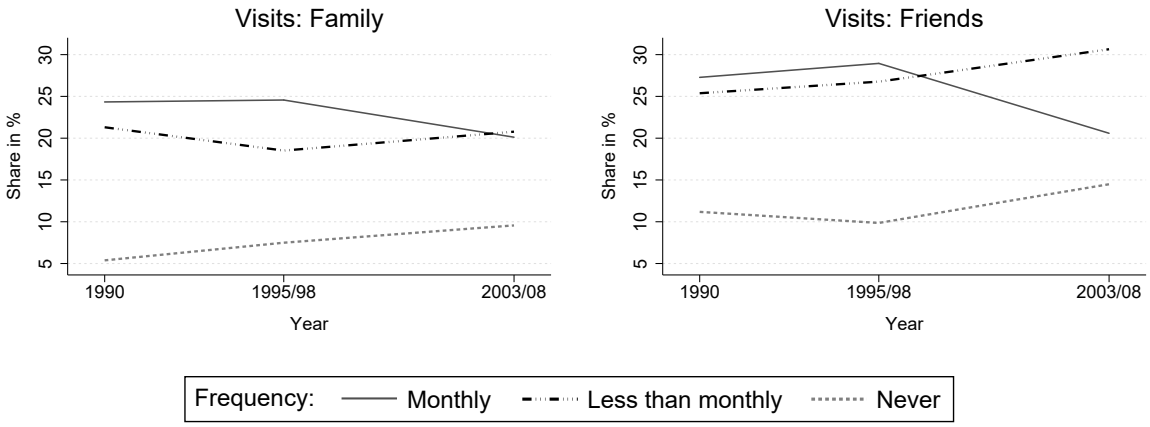

Figure 2.7. Time Trends of the Share of Elderly with Mutual Visits with Family and Friends by Frequency of Visit

Notes: These figures depict time trends of the share of elderly with mutual visits with (i) family and (ii) friends by frequency of visit. Due to the small sample sizes, a distinction by time to death is not meaningful. Therefore, shares refer to all elderly who were within five years of death in the given years. Shares do not add to 100% as the category "weekly visits" is not shown.
Sources: SOEPv30 (1984-2013) and GBE (2016), own calculations.

from reduced mobility because mobility strongly declines with age among the elderly (Ferrucci et al. 2016).

Figure 2.7 shows that the frequency of mutual visits with family and friends strongly decreased over time for West German elderly who were within their last five years of life.[37] Between 1990 and 2008, the shares of elderly with less than monthly mutual visits (only for mutual visits with friends) and without any visits (for both types of mutual visits) increased by more than five percentage points, respectively. Thus, it is highly likely that increased social isolation among elderly before death contributed to the decline in end-of-life life satisfaction over time. Table 2D.2 in Appendix 2D illustrates that this was indeed the case. Upon controlling for the frequency of mutual visits with family and friends in individual-level life satisfaction regressions, the year dummy coefficient estimates decrease in absolute value, which suggests that the decline in end-of-life life satisfaction relative to the baseline year 1990 became smaller. Table 2D.3 in Appendix 2D shows that this result also holds for individual-level life satisfaction regressions by time-to-death, although most of the coefficients are lacking statistical significance due to the small sample size. Overall,

[37]Information on these indicators was only collected for five years between 1985 and 2010. Therefore, in this figure, I pool the data of two years together (if applicable) and do not distinguish by time to death.

these results suggest that increased social isolation among West German elderly is an important explanation for the decline in end-of-life life satisfaction over time.[38]

2.7 Conclusion and Discussion

Given the strong increase in human life expectancy throughout the last decades, this study asks: Are longer lives more satisfied lives? Using data of the German Socio-Economic Panel, this study suggests that the answer might be no. The primary reason for this conclusion is that increases in lifetime were accompanied by substantial losses of life satisfaction in old age. Between 1985 and 2010, average life satisfaction of West German elderly who were in their last five years of life strongly declined. Moreover, the proportion of expected satisfied lifetime to expected total lifetime at age 60 decreased since 1985. I show that both results are robust to a large set of alternative explanations, thereby supporting the conclusion that increased longevity led to the decline of life satisfaction in old age. Two important mechanisms that contribute to this decline in life satisfaction are health and social isolation. Evidence on the former mechanism supports the expansion of morbidity hypothesis. Evidence on the latter mechanism stresses the importance of social integration of the elderly in aging societies.

Should people and policymakers further invest in life extensions? In light of the rather pessimistic results, the answer to this question merits discussion that is more thorough. Although this study finds that life satisfaction in old age substantially declined with increased longevity, it should not be ignored that roughly half of the additional lifetime that West Germans gained between 1985 and 2010 was, on average, spent in satisfaction in 2010. This increase in satisfied lifetime, can justify further investments in life-extending technologies and policies. However, it is important to complement these investments with investments that improve the quality of life in old age. This study clearly shows that overall life satisfaction, on average, decreases with improved longevity if decision makers invest too little in quality-of-life-improving policies. Therefore, in the future decision makers should invest more into policies that improve the quality of human life. Contrary to life-extending policies, quality-of-life improving policies may have a more positive effect on increasing overall life satisfaction. This is because quality-of-life improving policies

[38] I investigated three additional indicators of social isolation: single household status, partnership status, and widowhood. I find that these indicators did not contribute to the decline in end-of-life life satisfaction of West German elderly between 1985 and 2010 (results available upon request).

would increase satisfaction during a person's lifetime, and furthermore, may also extend the length of life itself as more satisfied people tend to live longer.

Which quality-of-life improving policies are highly likely to ensure that increases in longevity come along with increases in life satisfaction? As suggested by the analysis of potential mechanisms, potential candidates would be a subsidization of quality-of-life improving health research or policies that aim at a better integration of the elderly in today's societies (e.g., policies that improve the mobility of elderly). However, these are just some possible examples of policy options, and further research on these and other potential mechanisms is required to decide upon the policies that are most promising. Similarly, future research needs to explore potential heterogeneity in order to narrow down the group of most important recipients. The latter will also be important in the context of current euthanasia debates as it helps to identify groups of people for whom euthanasia laws should be relaxed in the future.

References

Ahlert, M., F. Breyer, and L. Schwettmann (2016): "How You Ask Is What You Get: Framing Effects in Willingness-to-Pay for a QALY," *Social Science & Medicine*, 150, 40–48.

Baetschmann, G. (2014): "Heterogeneity in the Relationship Between Happiness and Age: Evidence from the German Socio-Economic Panel," *German Economic Review*, 15, 393–410.

Baird, B. M., R. E. Lucas, and M. B. Donnellan (2010): "Life Satisfaction Across the Lifespan: Findings from Two Nationally Representative Panel Studies," *Social Indicators Research*, 99, 183–203.

Benjamin, D. J., O. Heffetz, M. S. Kimball, and A. Rees-Jones (2012): "What Do You Think Would Make You Happier? What Do You Think You Would Choose?" *American Economic Review*, 102, 2083–2110.

Benjamin, D. J., O. Heffetz, M. S. Kimball, and N. Szembrot (2014): "Beyond Happiness and Satisfaction: Toward Well-Being Indices Based on Stated Preference," *American Economic Review*, 104, 2698–2735.

Berg, A. I., L. Hassing, V. Thorvaldsson, and B. Johansson (2011): "Personality and Personal Control Make a Difference for Life Satisfaction in the Oldest-Old: Findings in a Longitudinal Population-Based Study of Individuals 80 and Older," *European Journal of Ageing*, 8, 13–20.

Blanchflower, D. G. and A. J. Oswald (2004): "Well-Being over Time in Britain and the USA," *Journal of Public Economics*, 88, 1359–1386.

——— (2008): "Is Well-Being U-Shaped over the Life Cycle?" *Social Science & Medicine*, 66, 1733–1749.

——— (2017): "Do Humans Suffer a Psychological Low in Midlife? Two Approaches (With and Without Controls) in Seven Data Sets." NBER Working Papers 23724, National Bureau of Economic Research.

Burkhauser, R. V., P. Giles, D. R. Lillard, and J. Schwarze (2005): "Until Death Do Us Part: An Analysis of the Economic Well-Being of Widows in Four Countries," *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 60, S238–S246.

Case, A. and A. Deaton (2015): "Rising Morbidity and Mortality in Midlife Among White Non-Hispanic Americans in the 21st Century," *Proceedings of the National Academy of Sciences*, 112, 15078–15083.

Chappell, N. L. and M. Badger (1989): "Social Isolation and Well-Being," *Journal of Gerontology*, 44, S169–S176.

Chatterji, S., J. Byles, D. Cutler, T. Seeman, and E. Verdes (2015): "Health, Functioning, and Disability in Older Adults – Present Status and Future Implications," *Lancet*, 385, 563–575.

Chetty, R., M. Stepner, S. Abraham, S. Lin, B. Scuderi, N. Turner, A. Bergeron, and D. Cutler (2016): "The Association Between Income and Life Expectancy in the United States, 2001-2014," *Journal of the American Medical Association*, 315, 1750–1766.

Chiang, C. L. (1960): "A Stochastic Study of the Life Table and Its Applications: I. Probability Distributions of the Biometric Functions," *Biometrics*, 16, 618–635.

——— (1984): *The Life Table and Its Applications*, Malabar, FL: Krieger Publishing Company.

Clark, A. E., P. Frijters, and M. A. Shields (2008): "Relative Income, Happiness, and Utility: An Explanation for the Easterlin Paradox and Other Puzzles," *Journal of Economic Literature*, 46, 95–144.

Clark, A. E. and A. J. Oswald (1994): "Unhappiness and Unemployment," *Economic Journal*, 104, 648–659.

Crimmins, E. M. and H. Beltrán-Sánchez (2011): "Mortality and Morbidity Trends: Is There Compression of Morbidity?" *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 66B, 75–86.

Crimmins, E. M., M. D. Hayward, A. Hagedorn, Y. Saito, and N. Brouard (2009): "Change in Disability-Free Life Expectancy for Americans 70 Years Old and Older," *Demography*, 46, 627–646.

Cutler, D. M., K. Ghosh, and M. B. Landrum (2014): "Evidence for Significant Compression of Morbidity in the Elderly U.S. Population," in *Discoveries in the Economics of Aging*, ed. by D. A. Wise, London and Chicago: The University of Chicago Press, 21–31.

Deaton, A. (2008): "Income, Health, and Well-Being Around the World: Evidence from the Gallup World Poll," *Journal of Economic Perspectives*, 22, 53–72.

DePinho, R. A. (2000): "The Age of Cancer," *Nature*, 408, 248–254.

Di Tella, R., R. J. MacCulloch, and A. J. Oswald (2001): "Preferences over Inflation and Unemployment: Evidence from Surveys of Happiness," *American Economic Review*, 91, 335–341.

Diener, E., E. M. Suh, R. E. Lucas, and H. L. Smith (1999): "Subjective Well-Being: Three Decades of Progress," *Psychological Bulletin*, 125, 276–302.

Ditto, P. H., J. A. Druley, K. A. Moore, J. H. Danks, and W. D. Smucker (1996): "Fates Worse Than Death: The Role of Valued Life Activities in Health-State Evaluations," *Health Psychology*, 15, 332–343.

Easterlin, R. A. (2002): *Happiness and Economics*, Cheltenham: Edward Elgar.

——— (2003): "Explaining Happiness," *Proceedings of the National Academy of Sciences*, 100, 11176–11183.

Engelman, M., V. Canudas-Romo, and E. M. Agree (2010): "The Implications of Increased Survivorship for Mortality Variation in Aging Populations," *Population and Development Review*, 36, 511–539.

Ferrucci, L., R. Cooper, M. Shardell, E. M. Simonsick, J. A. Schrack, and D. Kuh (2016): "Age-Related Change in Mobility: Perspectives from Life Course Epidemiology and Geroscience," *Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, 71, 1184–1194.

Fischer, B., H. Telser, and P. Zweifel (2017): "End-of-Life Healthcare Expenditure: Testing Economic Explanations Using a Discrete Choice Experiment," Working Papers 201701, Polynomics.

Fleurbaey, M. and H. Schwandt (2015): "Do People Seek to Maximize Their Subjective Well-Being?" IZA Discussion Papers 9450, Institute for the Study of Labor (IZA).

Frey, B. S. and A. Stutzer (2002): "What Can Economists Learn from Happiness Research?" *Journal of Economic Literature*, 40, 402–435.

Fries, J. F. (1980): "Aging, Natural Death, and the Compression of Morbidity," *New England Journal of Medicine*, 303, 130–135.

Frijters, P. and T. Beatton (2012): "The Mystery of the U-Shaped Relationship Between Happiness and Age," *Journal of Economic Behavior & Organization*, 82, 525 – 542.

Frijters, P., J. P. Haisken-DeNew, and M. A. Shields (2004): "Money Does Matter! Evidence from Increasing Real Income and Life Satisfaction in East Germany Following Reunification," *American Economic Review*, 94, 730–740.

German Statistical Office (2012a): "Periodensterbetafeln für Deutschland. Allgemeine Sterbetafeln, abgekürzte Sterbetafeln und Sterbetafeln, 1871/1881 bis 2008/2010," Wiesbaden. www.destatis.de/DE/Publikationen/Thematisch/Bevoelkerung/Bevoelkerungsbewegung/PeriodensterbetafelnPDF_5126202.pdf?__blob=publicationFile (Accessed on 29 February 2016).

——— (2012b): "Periodensterbetafeln für Deutschland. Früheres Bundesgebiet, neue Länder sowie die Bundesländer, 2009/2011," Wiesbaden. www.destatis.de/DE/Publikationen/Thematisch/Bevoelkerung/Bevoelkerungsbewegung/PeriodensterbetafelnBundeslaender5126204117004.pdf?__blob=publicationFile (Accessed on 29 February 2016).

——— (2016a): "Gestorbene: Deutschland, Jahre, Geschlecht, Altersjahre (1970-2014)," Wiesbaden. https://www-genesis.destatis.de/genesis/online/data;jsessionid=B918F237636C71F76C7B116A80DC6351.tomcat_GO_2_1?operation=abruftabelleAbrufen&selectionname=12613-0003&levelindex=0&levelid=1470408175298&index=3 (Accessed on 08 August 2016).

——— (2016b): "Zusammenleben von Generationen," Wiesbaden. https://www.destatis.de/DE/PresseService/Presse/Pressekonferenzen/2016/zusammenleben_generationen/zusammenleben_generationen_uebersicht.html (Accessed on 15 January 2018).

Gerstorf, D., C. A. Hoppmann, C. E. Löckenhoff, F. J. Infurna, J. Schupp, G. G. Wagner, and N. Ram (2016): "Terminal Decline in Well-Being: The Role of Social Orientation," *Psychology and Aging*, 31, 149–165.

Gerstorf, D., N. Ram, R. Estabrook, J. Schupp, G. G. Wagner, and U. Lindenberger (2008a): "Life Satisfaction Shows Terminal Decline in Old Age: Longitudinal Evidence from the German Socio-Economic Panel Study (SOEP)," *Developmental Psychology*, 44, 1148–1159.

Gerstorf, D., N. Ram, G. Mayraz, M. Hidajat, U. Lindenberger, G. G. Wagner, and J. Schupp (2010): "Late-Life Decline in Well-Being Across Adulthood in Germany, the UK, and the US: Something Is Seriously Wrong at the End of Life," *Psychology and Aging*, 25, 477–485.

Gerstorf, D., N. Ram, and C. Röcke (2008b): "Decline in Life Satisfaction in Old Age: Longitudinal Evidence for Links to Distance-to-Death," *Psychology and Aging*, 23, 154–168.

Gesundheitsberichterstattung des Bundes (GBE) (2016): "Sterbefälle, Sterbeziffern je 100000 Einwohner (altersstandardisiert) (ab 1980). Gliederungsmerkmale: Jahre, Region, Alter, Geschlecht, Nationalität, Todesursachen laut Europäischer Kurzliste, Art der Standardisierung," `www.gbe-bund.de/stichworte/Todesursachenstatistik.html` (Accessed on 14 June 2016).

Gruenberg, E. M. (1977): "The Failures of Success," *Milbank Memorial Fund Quarterly. Health and Society*, 55, 3–24.

Gwozdz, W. and A. Sousa-Poza (2010): "Ageing, Health and Life Satisfaction of the Oldest Old: An Analysis for Germany," *Social Indicators Research*, 97, 397–417.

Helliwell, J. F. (2003): "How's Life? Combining Individual and National Variables to Explain Subjective Well-Being," *Economic Modelling*, 20, 331–360.

——— (2006): "Well-Being, Social Capital and Public Policy: What's New?" *Economic Journal*, 116, C34–C45.

Imai, K. and S. Soneji (2007): "On the Estimation of Disability-Free Life Expectancy: Sullivan's Method and Its Extension," *Journal of the American Statistical Association*, 102, 1199–1211.

Jagger, C. and J.-M. Robine (2011): "Healthy Life Expectancy," in *International Handbook of Adult Mortality*, ed. by G. R. Rogers and M. E. Crimmins, Dordrecht: Springer Netherlands, 551–568.

Jagger, C., H. Van Oyen, and J.-M. Robine (2014): "Health Expectancy Calculation by the Sullivan Method: A Practical Guide (4th ed.)," Technical Report, Institute for Ageing, Newcastle University.

Kassenboehmer, S. C. and J. P. Haisken-DeNew (2009): "You're Fired! The Causal Negative Effect of Entry Unemployment on Life Satisfaction," *Economic Journal*, 119, 448–462.

——— (2012): "Heresy or Enlightenment? The Well-Being Age U-Shape Effect is Flat," *Economics Letters*, 117, 235–238.

Klein, T. (1996): "Determinants of Institutionalization in Old Age," in *Long-Term Care: Economic Issues and Policy Solutions*, ed. by R. Eisen and F. A. Sloan, Boston, MA: Springer US, 103–113.

Kroh, M. (2014): "Documentation of Sample Sizes and Panel Attrition in the German Socio Economic Panel, SOEP (1984 until 2012)," SOEP Survey Papers (Series D) 177, Deutsches Institut für Wirtschaftsforschung (DIW Berlin).

Kroh, M., R. Pischner, M. Spiess, and G. G. Wagner (2008): "On the Treatment of Non-Original Sample Members in the German Household Panel Study (SOEP): Tracing and Weighting," *Methoden–Daten–Analysen*, 2, 179–198.

Kunst, A. E., P. T. Okma-Keulen, and R. Veenhoven (1994): "Happy Life Expectancy in 5 European Countries," Paper presented at the XIII World Congress of Sociology, Bielefeld, Germany, 18–23 July 1994.

Kunzmann, U., T. D. Little, and J. Smith (2000): "Is Age-Related Stability of Subjective Well-Being a Paradox? Cross-Sectional and Longitudinal Evidence from the Berlin Aging Study," *Psychology and Aging*, 15, 511–526.

Lindgren, B. (2016): "The Rise in Life Expectancy, Health Trends Among the Elderly, and the Demand for Health and Social Care," NIER Working Papers 142, National Institute of Economic Research (NIER).

Mahne, K. and O. Huxhold (2017): "Nähe auf Distanz: Bleiben die Beziehungen zwischen älteren Eltern und ihren erwachsenen Kindern trotz wachsender Wohnentfernungen gut?" in *Altern im Wandel: Zwei Jahrzehnte Deutscher Alterssurvey (DEAS)*, ed. by K. Mahne, J. K. Wolff, J. Simonson, and C. Tesch-Römer, Wiesbaden: Springer VS, 215–230.

Manton, K. G. (1982): "Changing Concepts of Morbidity and Mortality in the Elderly Population," *Milbank Memorial Fund Quarterly. Health and Society*, 60, 183–244.

Mathers, C. D. (1991): "Health Expectancies in Australia 1981 and 1988," Australian Institute of Health: AGPS, Canberry.

Mathers, C. D. and J.-M. Robine (1997): "How Good Is Sullivan's Method for Monitoring Changes in Population Health Expectancies," *Journal of Epidemiology and Community Health*, 51, 80–86.

McNamee, P. and S. Mendolia (2014): "The Effect of Chronic Pain on Life Satisfaction: Evidence from Australian Data," *Social Science & Medicine*, 121, 65–73.

Newman, S. C. (1988): "A Markov Process Interpretation of Sullivan's Index of Morbidity and Mortality," *Statistics in Medicine*, 7, 787–794.

OECD (2016): "Life Expectancy at Birth (Indicator)," doi:`http://dx.doi.org/10.1787/27e0fc9d-en` (Accessed on 15 July 2016).

Olshansky, S., M. Rudberg, B. Carnes, C. Cassel, and J. Brody (1991): "Trading Off Longer Life for Worsening Health: The Expansion of Morbidity Hypothesis," *Journal of Aging and Health*, 3, 194–216.

Olshansky, S. J., T. Antonucci, L. Berkman, R. H. Binstock, A. Boersch-Supan, J. T. Cacioppo, B. A. Carnes, L. L. Carstensen, L. P. Fried, D. P. Goldman, J. Jackson, M. Kohli, J. Rother, Y. Zheng, and J. Rowe (2012): "Differences in Life Expectancy due to Race and Educational Differences Are Widening, and Many May Not Catch Up," *Health Affairs*, 31, 1803–1813.

Oreopoulos, P. (2007): "Do Dropouts Drop Out Too Soon? Wealth, Health and Happiness from Compulsory Schooling," *Journal of Public Economics*, 91, 2213–2229.

Oswald, A. J. (1997): "Happiness and Economic Performance," *Economic Journal*, 107, 1815–1831.

Oswald, A. J. and N. Powdthavee (2008): "Does Happiness Adapt? A Longitudinal Study of Disability with Implications for Economists and Judges," *Journal of Public Economics*, 92, 1061–1077.

Palgi, Y., A. Shrira, M. Ben-Ezra, T. Spalter, D. Shmotkin, and G. Kav (2010): "Delineating Terminal Change in Subjective Well-Being and Subjective Health: Brief Report," *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 65B, 61–64.

Pennington, M., R. Baker, W. Brouwer, H. Mason, D. G. Hansen, A. Robinson, C. Donaldson, and the EuroVaQ Team (2015): "Comparing WTP Values of Different Types of QALY Gain Elicited from the General Public," *Health Economics*, 24, 280–293.

Perenboom, R. J. M., L. M. Van Herten, H. C. Boshuizen, and G. A. M. Van Den Bos (2004): "Trends in Life Expectancy in Wellbeing," *Social Indicators Research*, 65, 227–244.

Pinquart, M. and S. Sörensen (2000): "Influences of Socioeconomic Status, Social Network, and Competence on Subjective Well-Being in Later Life: A Meta-Analysis," *Psychology and Aging*, 15, 187–224.

Powdthavee, N. (2008): "Putting a Price Tag on Friends, Relatives, and Neighbours: Using Surveys of Life Satisfaction to Value Social Relationships," *Journal of Socio-Economics*, 37, 1459–1480.

Rosenthal, R. A. and S. M. Kavic (2004): "Assessment and Management of the Geriatric Patient," *Critical Care Medicine*, 32, S92–S105.

Rubin, E. B., A. E. Buehler, and S. D. Halpern (2016): "States Worse Than Death Among Hospitalized Patients With Serious Illnesses," *JAMA Internal Medicine*, 176, 1557–1559.

Schilling, O. (2006): "Development of Life Satisfaction in Old Age: Another View on the 'Paradox'," *Social Indicators Research*, 75, 241–271.

Schnell, R. and M. Trappmann (2006): "Konsequenzen der Panelmortalität im SOEP für Schätzungen der Lebenserwartung," in *Stichprobenqualität in Bevölkerungsumfragen*, ed. by F. Faulbaum and C. Wolf, Bonn: Informationszentrum Sozialwissenschaften, 139–158.

Schwandt, H. (2016): "Unmet Aspirations as an Explanation for the Age U-Shape in Wellbeing," *Journal of Economic Behavior & Organization*, 122, 75–87.

Seshamani, M. and A. Gray (2004): "Ageing and Health-Care Expenditure: The Red Herring Argument Revisited," *Health Economics*, 13, 303–314.

Steffel, M. and D. M. Oppenheimer (2009): "Happy by What Standard? The Role of Interpersonal and Intrapersonal Comparisons in Ratings of Happiness," *Social Indicators Research*, 92, 69–79.

Steptoe, A., A. Deaton, and A. A. Stone (2015): "Subjective Wellbeing, Health, and Ageing," *Lancet*, 385, 640–648.

Sullivan, D. F. (1971): "A Single Index of Mortality and Morbidity," *HSMHA Health Reports*, 86, 347–354.

Veenhoven, R. (1996): "Happy Life-Expectancy. A Comprehensive Measure of Quality-of-Life in Nations," *Social Indicators Research*, 39, 1–58.

——— (2008): "Healthy Happiness: Effects of Happiness on Physical Health and the Consequences for Preventive Health Care," *Journal of Happiness Studies*, 9, 449–469.

Vogel, N., D. Gerstorf, N. Ram, J. Goebel, and G. G. Wagner (2017): "Terminal Decline in Well-Being Differs Between Residents in East Germany and West Germany," *International Journal of Behavioral Development*, 41, 115–126.

Wagner, G. G., J. R. Frick, and J. Schupp (2007): "The German Socio-Economic Panel Study (SOEP) – Scope, Evolution and Enhancements," *Schmollers Jahrbuch*, 127, 139–169.

Winkelmann, L. and R. Winkelmann (1995): "Happiness and Unemployment: A Panel Data Analysis for Germany," *Konjunkturpolitik*, 41, 293–307.

Wunder, C., A. Wiencierz, J. Schwarze, and H. Kchenhoff (2013): "Well-Being over the Life Span: Semiparametric Evidence from British and German Longitudinal Data," *Review of Economics and Statistics*, 95, 154–167.

Yang, Y. (2008): "Long and Happy Living: Trends and Patterns of Happy Life Expectancy in the U.S., 1970-2000," *Social Science Research*, 37, 1235–1252.

Yang, Y. and M. Waliji (2010): "Increment–Decrement Life Table Estimates of Happy Life Expectancy for the U.S. Population," *Population Research and Policy Review*, 29, 775–795.

Zweifel, P., S. Felder, and M. Meiers (1999): "Ageing of Population and Health Care Expenditure: A Red Herring?" *Health Economics*, 8, 485–496.

Appendix

2A Time-to-Death Approach: Supplementary Tables and Figures

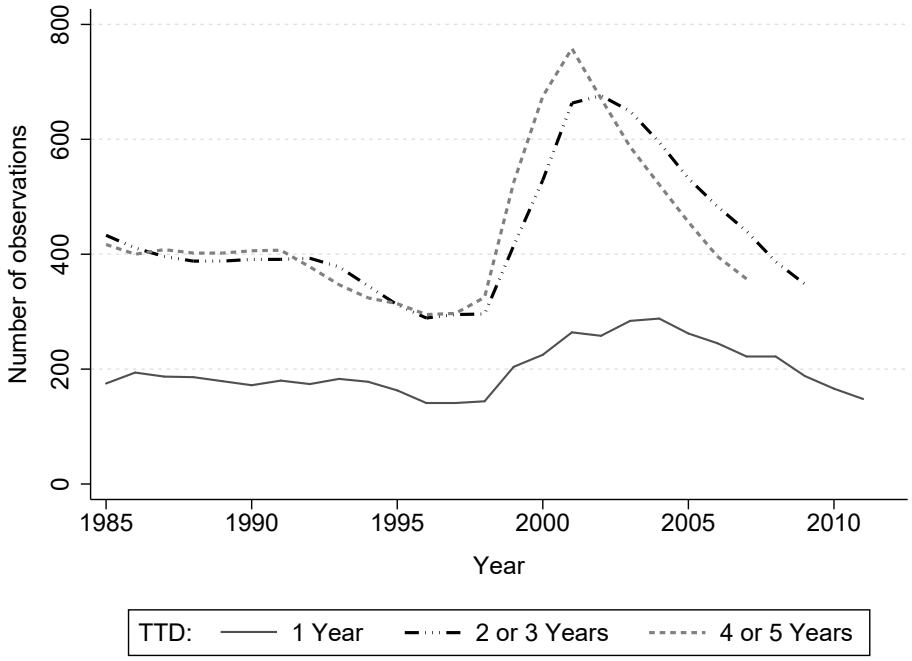

Figure 2A.1. Number of Observations by Time to Death and Year

Notes: TTD = time to death. Three-year pooled average data set. For the main variables (life satisfaction and age at death), estimates rely on 26,264 person-year-observations, which are obtained from 2,446 deceased respondents. The evolution of the number of observations for these variables is depicted in this figure. Between 1998 and 2010, the number of observations is higher because of two refreshment samples, which were added in 1998 and 2000.
Source: SOEPv30 (1984-2013) and GBE (2016), own calculations.

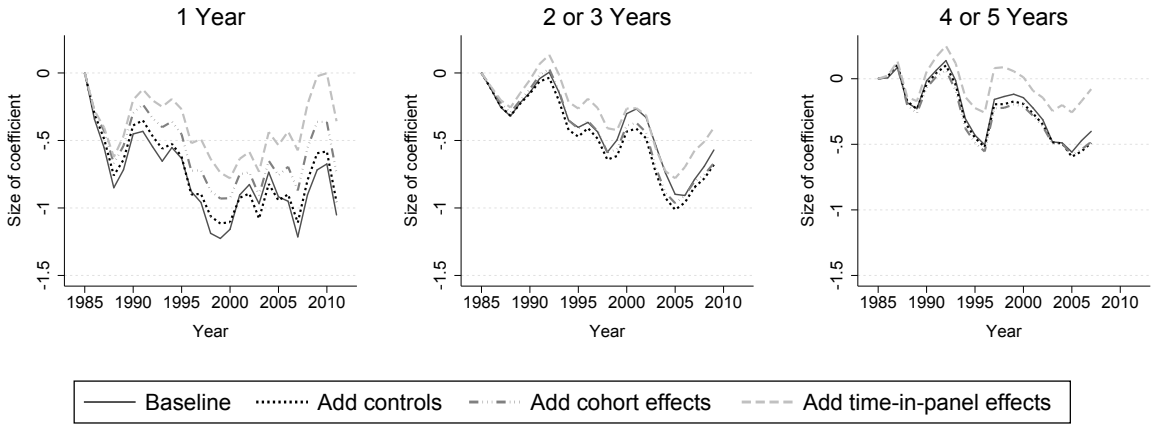

Figure 2A.2. Role of Compositional Effects, Cohort Effects, and Time-in-Panel Effects: Year Dummy Coefficient Estimates

Notes: These figures illustrate the change of the year dummy coefficient estimates after sequentially adding controls for compositional effects, cohort effects, and time-in-panel effects to weighted individual-level life satisfaction regressions, which regress life satisfaction on a set of year dummies. Separate regressions were estimated by time to death. Coefficient estimates are relative to the year 1985, which is the base year in all regressions. Regression results are reported in Table 2A.1 in Appendix 2A.
Sources: SOEPv30 (1984-2013) and GBE (2016), own calculations.

Table 2A.1. Compositional Effects, Cohort Effects, and Time-in-Panel Effects: Life Satisfaction Regressions by Time to Death

	1 Year				2 or 3 Years				4 or 5 Years			
	Baseline	Add Controls	Add Cohorts	Add TiP	Baseline	Add Controls	Add Cohorts	Add TiP	Baseline	Add Controls	Add Cohorts	Add TiP
Constant	6.481***	5.156***	5.378***	5.509***	6.751***	4.994***	5.331***	5.426***	6.866***	4.985***	4.392***	4.550***
	(0.245)	(0.555)	(0.835)	(0.836)	(0.275)	(0.501)	(0.716)	(0.712)	(0.184)	(0.377)	(0.879)	(0.908)
1986	-0.342*	-0.301	-0.284*	-0.279	-0.120	-0.119	-0.108	-0.103	0.007	0.018	0.013	0.024
	(0.195)	(0.184)	(0.172)	(0.172)	(0.115)	(0.110)	(0.106)	(0.106)	(0.093)	(0.091)	(0.091)	(0.092)
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1997	-0.959***	-0.896***	-0.721**	-0.490	-0.437	-0.487	-0.432	-0.260	-0.158	-0.195	-0.225	0.081
	(0.331)	(0.336)	(0.343)	(0.356)	(0.331)	(0.328)	(0.337)	(0.344)	(0.241)	(0.245)	(0.249)	(0.260)
1998	-1.188***	-1.061***	-0.874**	-0.637*	-0.587*	-0.642**	-0.587*	-0.409	-0.136	-0.194	-0.222	0.088
	(0.341)	(0.344)	(0.357)	(0.373)	(0.329)	(0.327)	(0.337)	(0.344)	(0.234)	(0.236)	(0.244)	(0.253)
1999	-1.226***	-1.115***	-0.930***	-0.738**	-0.496	-0.615**	-0.567*	-0.423	-0.118	-0.173	-0.196	0.057
	(0.323)	(0.334)	(0.349)	(0.363)	(0.314)	(0.310)	(0.323)	(0.330)	(0.223)	(0.227)	(0.237)	(0.244)
2000	-1.158***	-1.105***	-0.928***	-0.780**	-0.301	-0.433	-0.389	-0.264	-0.144	-0.191	-0.211	0.011
	(0.314)	(0.331)	(0.348)	(0.356)	(0.306)	(0.301)	(0.314)	(0.320)	(0.216)	(0.222)	(0.235)	(0.239)
2001	-0.904***	-0.925***	-0.750**	-0.636*	-0.266	-0.412	-0.369	-0.261	-0.223	-0.265	-0.283	-0.092
	(0.300)	(0.320)	(0.341)	(0.346)	(0.300)	(0.296)	(0.312)	(0.317)	(0.208)	(0.221)	(0.233)	(0.236)
2002	-0.826***	-0.893***	-0.736**	-0.585*	-0.329	-0.484	-0.440	-0.313	-0.312	-0.348	-0.365	-0.144
	(0.297)	(0.324)	(0.348)	(0.354)	(0.297)	(0.295)	(0.314)	(0.319)	(0.214)	(0.228)	(0.241)	(0.244)
2003	-0.968***	-1.079***	-0.907***	-0.732**	-0.537*	-0.738**	-0.700**	-0.559*	-0.482**	-0.486**	-0.494**	-0.246
	(0.295)	(0.321)	(0.347)	(0.357)	(0.298)	(0.295)	(0.317)	(0.322)	(0.219)	(0.234)	(0.249)	(0.256)
2004	-0.735**	-0.822***	-0.650*	-0.438	-0.741**	-0.926***	-0.887***	-0.728**	-0.489**	-0.493**	-0.492*	-0.201
	(0.288)	(0.316)	(0.344)	(0.357)	(0.304)	(0.298)	(0.322)	(0.330)	(0.222)	(0.234)	(0.252)	(0.264)
2005	-0.917***	-0.944***	-0.758**	-0.535	-0.899***	-1.013***	-0.964***	-0.780**	-0.559**	-0.595**	-0.584**	-0.257
	(0.306)	(0.324)	(0.356)	(0.371)	(0.308)	(0.300)	(0.327)	(0.335)	(0.225)	(0.236)	(0.254)	(0.269)
2006	-0.950***	-0.899***	-0.695**	-0.433	-0.907***	-0.960***	-0.908***	-0.698**	-0.472**	-0.554**	-0.539**	-0.171
	(0.304)	(0.320)	(0.353)	(0.371)	(0.310)	(0.302)	(0.330)	(0.341)	(0.223)	(0.233)	(0.250)	(0.268)
2007	-1.214***	-1.109***	-0.870**	-0.573	-0.789**	-0.846***	-0.803**	-0.576*	-0.401*	-0.489**	-0.481*	-0.079
	(0.335)	(0.343)	(0.376)	(0.402)	(0.313)	(0.307)	(0.334)	(0.345)	(0.230)	(0.237)	(0.255)	(0.271)
2008	-0.903***	-0.795**	-0.554	-0.232	-0.692**	-0.782**	-0.754**	-0.510				
	(0.325)	(0.335)	(0.370)	(0.401)	(0.315)	(0.311)	(0.337)	(0.349)				
2009	-0.718**	-0.594*	-0.362	-0.024	-0.571*	-0.679**	-0.667**	-0.405				
	(0.343)	(0.360)	(0.393)	(0.423)	(0.315)	(0.310)	(0.336)	(0.351)				
2010	-0.674**	-0.581*	-0.363	-0.002								
	(0.326)	(0.352)	(0.391)	(0.416)								
2011	-1.051***	-0.957***	-0.744*	-0.357								
	(0.340)	(0.361)	(0.400)	(0.425)								
Controls	–	Yes	Yes	Yes	–	Yes	Yes	Yes	–	Yes	Yes	Yes
Cohort effects	–	–	Yes	Yes	–	–	Yes	Yes	–	–	Yes	Yes
TiP effects	–	–	–	Yes	–	–	–	Yes	–	–	–	Yes
Adj. R2	0.01	0.04	0.05	0.06	0.01	0.04	0.05	0.05	0.01	0.04	0.04	0.05
N	5,221	5,221	5,221	5,221	10,532	10,532	10,532	10,532	9,797	9,797	9,797	9,797

Notes: TiP = time-in-panel. Controls include an indicator for males, years of education, need-weighted deflated net household income, indicators for the interview month, and indicators for the state of residence. Cohort effects are controlled for by including five-year cohort dummies. Due to the small sample size for cohorts born before 1895 and after 1945, these cohorts are collapsed in a single dummy, respectively. Time-in-panel effects are modeled linearly. Standard errors that allow for correlation across observations of the same respondent are reported in parentheses. For a graphical representation of all year dummy coefficient estimates, see Figure 2A.2 in Appendix 2A.

* p<0.1 ** p<0.05 *** p<0.01.

Sources: SOEPv30 (1984-2013) and GBE (2016), own calculations.

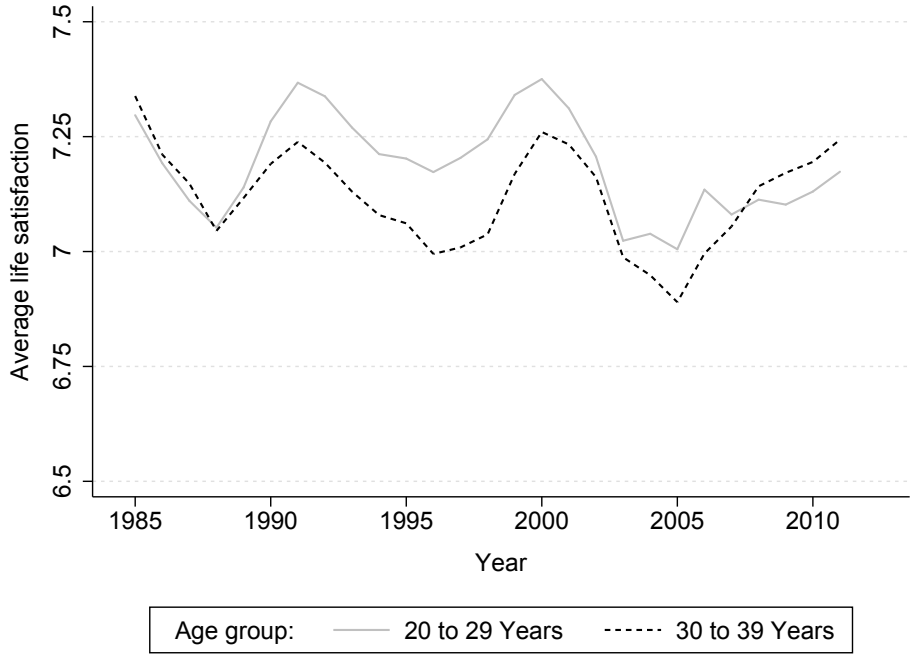

Figure 2A.3. Time Trends of Average Life Satisfaction by Age Group, Young West German Adults

Notes: Three-year averages for a given age group in a given year are estimated based on at least 2,557 respondent-year-observations. The focus is again on West Germans without migration backgrounds. Life satisfaction is measured on an eleven point Likert scale, ranging from 0 (very dissatisfied) to 10 (very satisfied).
Sources: SOEPv30 (1984-2013) and GBE (2016), own calculations.

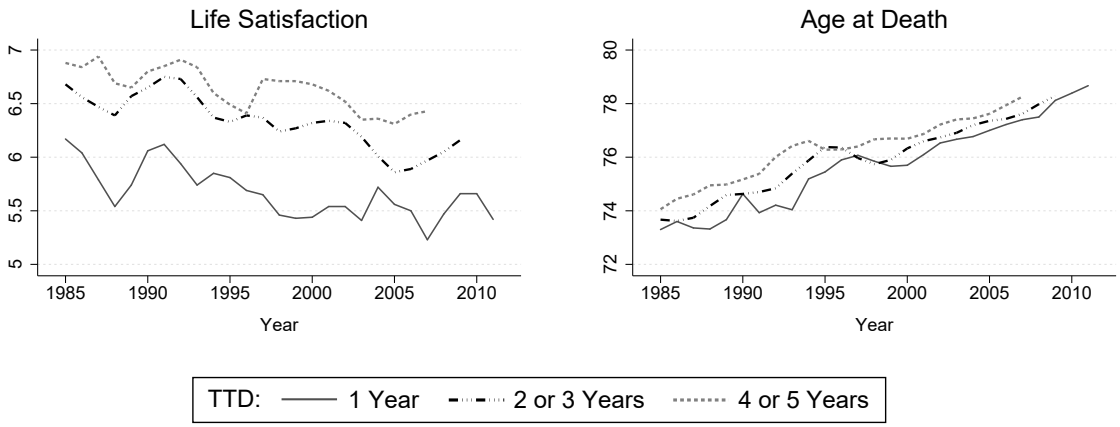

Figure 2A.4. Time Trends of Average Life Satisfaction and Average Age at Death by Time to Death, All West German Adults

Notes: TTD = time to death. Three-year averages are estimated based on 30,207 respondent-year-observations, which are obtained from 2,829 deceased respondents. The same sample restrictions as for the time-to-death sample apply, except that respondents who died at age 18 to 54 are also included in the sample. Life satisfaction is measured on an eleven point Likert scale, ranging from 0 (very dissatisfied) to 10 (very satisfied). Age at death is measured in years.
Sources: SOEPv30 (1984-2013) and GBE (2016), own calculations.

Figure 2A.5. Time Trends of Average Life Satisfaction and Average Age by Objective Death Probability

Notes: ODP = objective death probability, assigned from West German period life tables. Three-year averages are estimated based on 166,201 respondent-year-observations, which are obtained from 7,704 respondents. Estimates for a given objective death probability group in a given year are estimated based on at least 139 respondent-year-observations. The same sample restrictions as for the time-to-death sample apply, except that non-deceased respondents are also included in the sample. Life satisfaction is measured on an eleven point Likert scale, ranging from 0 (very dissatisfied) to 10 (very satisfied). Age is measured in years.
Sources: SOEPv30 (1984-2013) and German Statistical Office 2012a,b, own calculations.

Table 2A.2. Linear Probability Model for Study Dropout

	Dropout
Life satisfaction	-0.004*
	(0.002)
Post1995	-0.000
	(0.020)
Post1995 × life satisfaction	0.003
	(0.003)
2 or 3 years before death	0.036***
	(0.006)
Male	-0.009
	(0.006)
Age	0.001*
	(0.000)
Low education	-0.006
	(0.006)
Poor household	0.017**
	(0.008)
Single household	-0.023***
	(0.006)
Constant	0.029
	(0.030)
Adj. R2	0.01
N	7,383

Notes: Linear probability model estimates. The sample corresponds to the time-to-death sample, except that it excludes respondents who were one year before death. Low education is an indicator for those with less than 11 years of schooling. Poor household is an indicator for those living in households with net household income below the 60% poverty line. Standard errors that allow for correlation across observations of the same respondent are reported in parentheses.

* p<0.1 ** p<0.05 *** p<0.01.

Sources: SOEPv30 (1984-2013), own estimates.

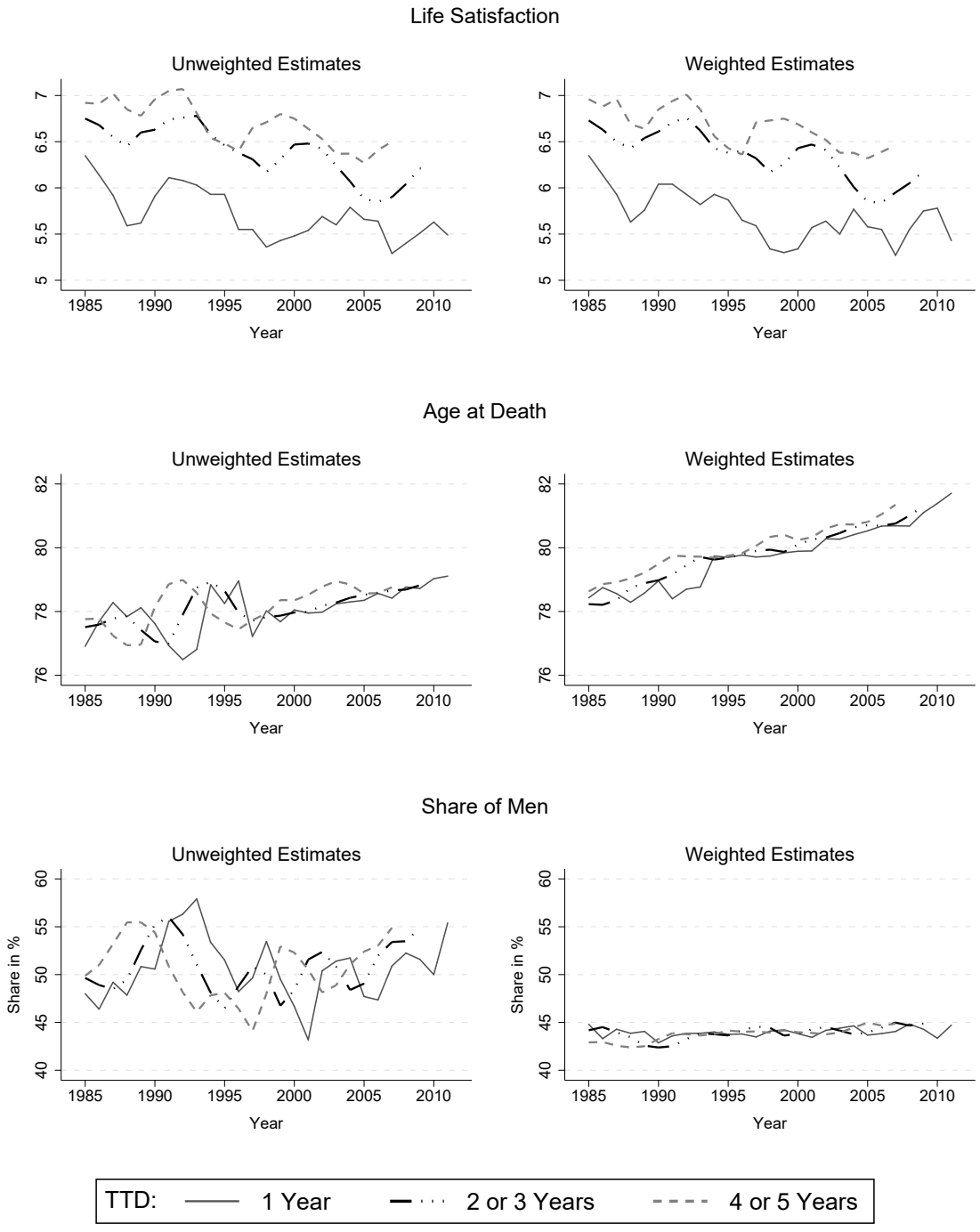

Figure 2A.6. Illustration of Weighting: Time Trends of Average Life Satisfaction, Average Age at Death, and the Share of Men by Time to Death

Notes: TTD = time to death. This graph illustrates the effect of weighting. Three-year averages are estimated based on 26,264 respondent-year-observations, which are obtained from 2,446 deceased respondents. Life satisfaction is measured on an eleven point Likert scale, ranging from 0 (very dissatisfied) to 10 (very satisfied). Age at death is measured in years.
Sources: SOEPv30 (1984-2013) and GBE (2016), own calculations.

2B Life-Expectancy Approach: Supplementary Figures

Figure 2B.1. Mortality Rates by Gender, Age, and Year

Sources: GBE (2016), own representation.

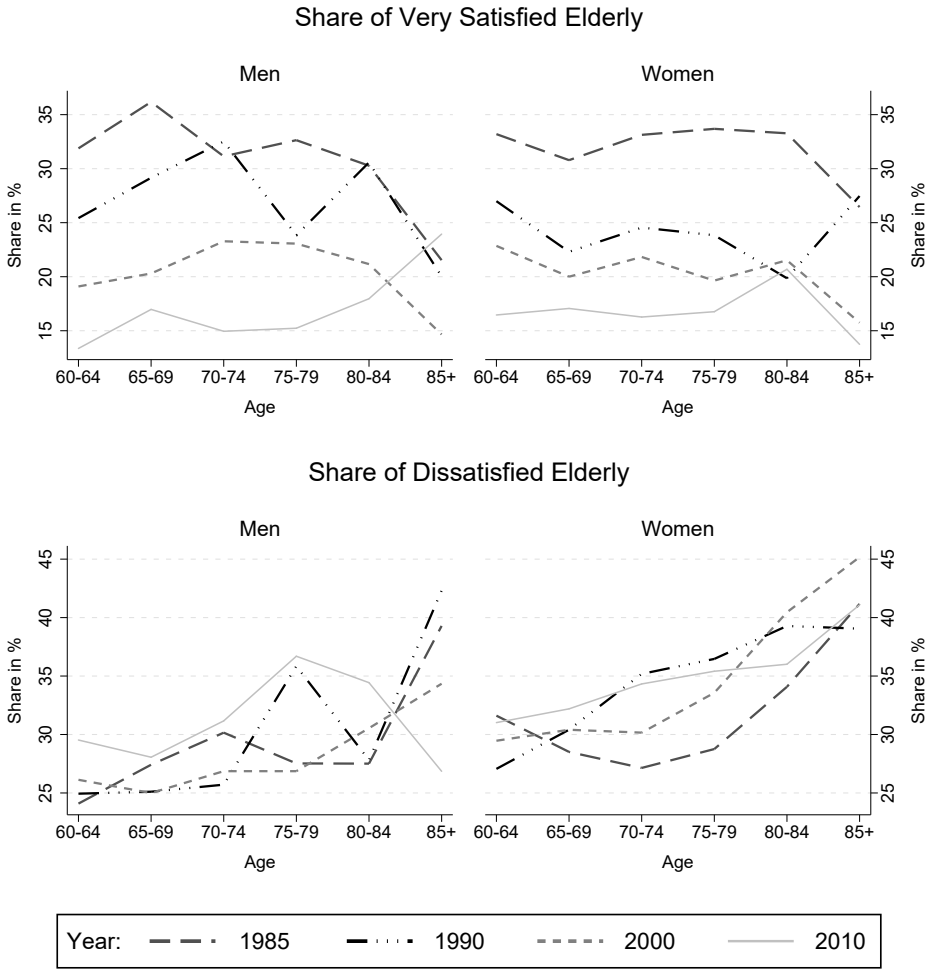

Figure 2B.2. Satisfaction Prevalence Rates by Gender, Age, and Year

Notes: Estimates represent three-year averages. Estimates are based on 11,852 (men) and 15,018 (women) respondent-year-observations. Respondents who indicated a life satisfaction score of 9 or 10 (0 to 6) were classified as very satisfied (dissatisfied).
Sources: SOEPv30 (1984-2011), own calculations.

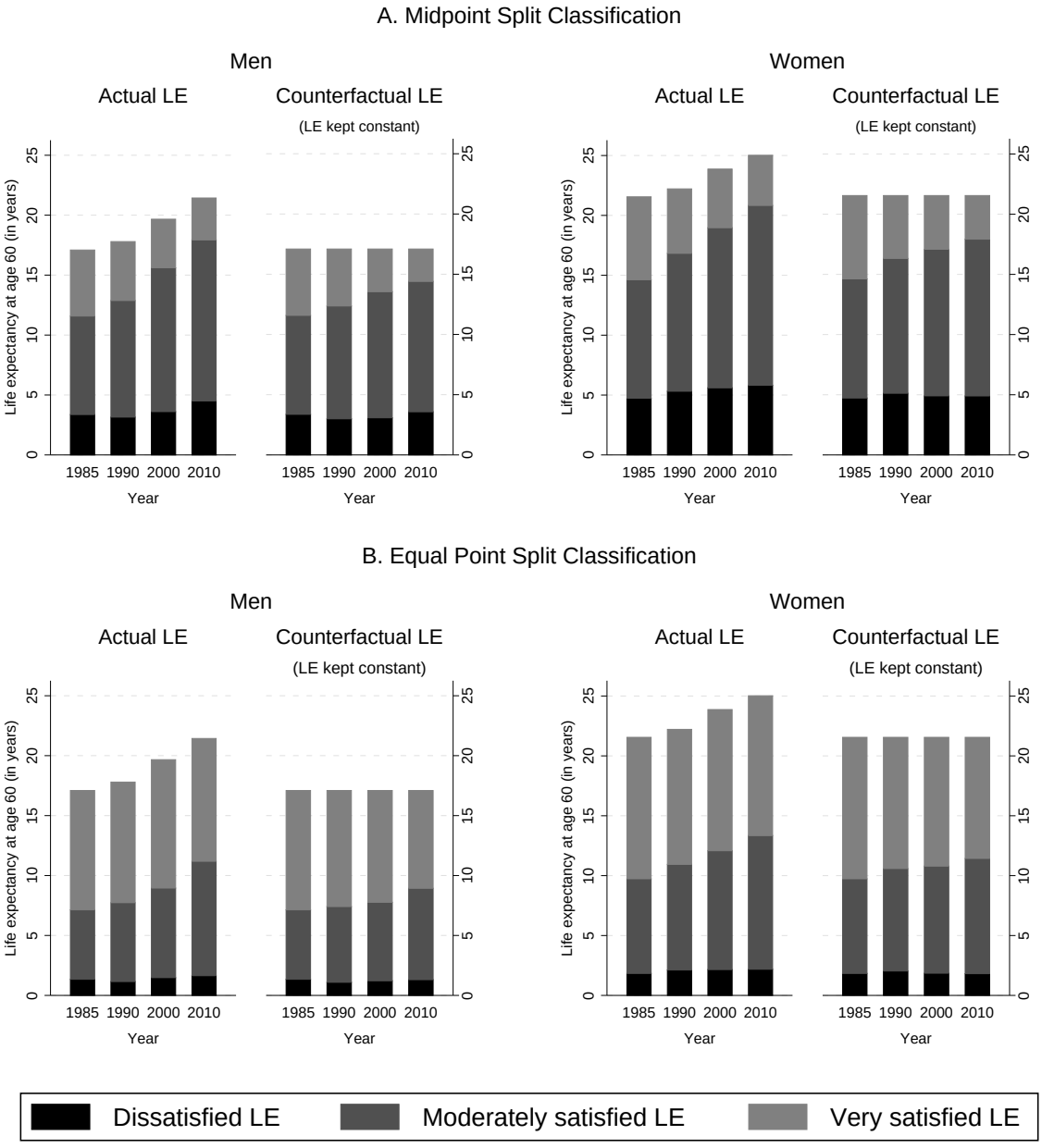

FIGURE 2B.3. ALTERNATIVE CLASSIFICATION RULES: ACTUAL AND COUNTERFACTUAL SATISFIED LIFE EXPECTANCY AT AGE 60 BY GENDER AND YEAR

Notes: LE = life expectancy. These figures show the evolution of total and satisfied life expectancy at age 60 in West Germany under alternative classification rules by gender. Under the midpoint split classification rule (the equal point split classification rule) respondents are classified as dissatisfied if their life satisfaction lies between 0 and 5 (0 and 4), as moderately satisfied if their life satisfaction lies between 6 and 8 (5 and 7), and as very satisfied if their life satisfaction lies between 9 and 10 (8 and 10). Figures on the left each illustrate changes of actual life expectancy (resulting from changes in both satisfaction prevalence *and* mortality), whereas figures on the right each illustrate changes of counterfactual life expectancy (resulting mainly from changes in satisfaction prevalence).
Sources: SOEPv30 (1984-2011) and German Statistical Office (2012a,b), own calculations.

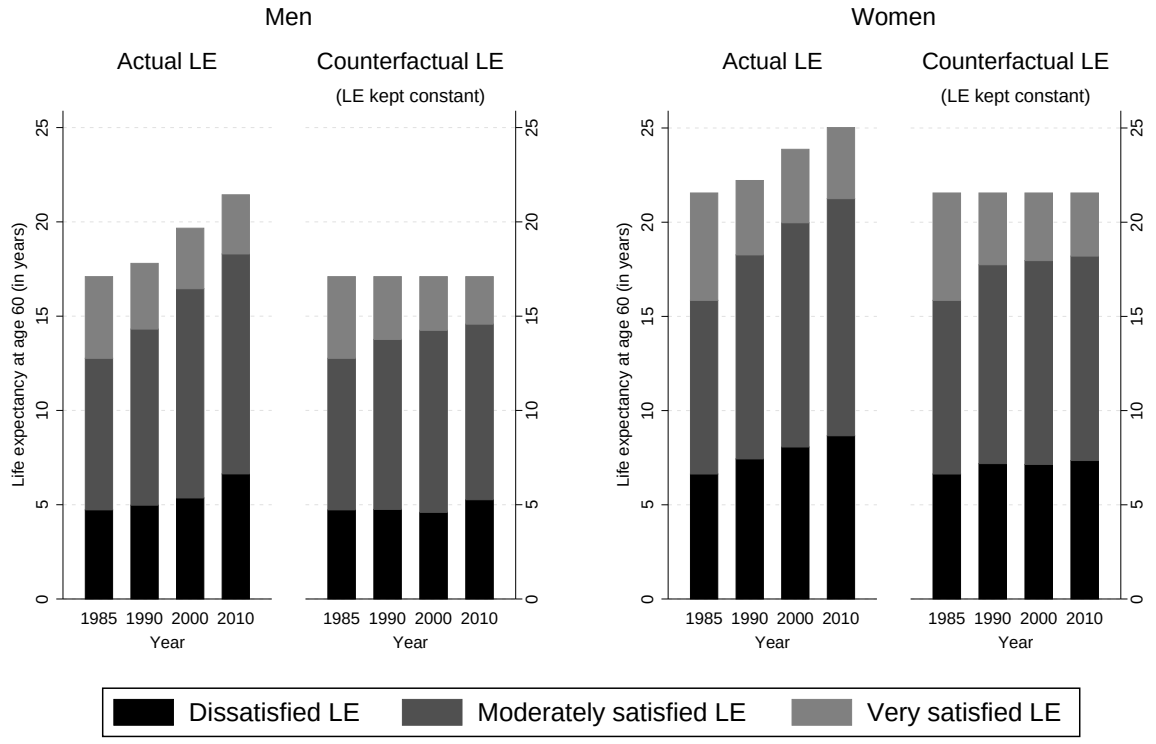

Figure 2B.4. Role of Cohort Effects: Actual and Counterfactual Satisfied Life Expectancy at Age 60 by Gender and Year

Notes: LE = life expectancy. These figures show the evolution of total and satisfied life expectancy at age 60 in West Germany by gender, after adjusting satisfaction prevalence rates for cohort effects in the estimations. To adjust for cohort effects, five-year cohort dummies were used. Figures on the left each illustrate changes of actual life expectancy (resulting from changes in both satisfaction prevalence *and* mortality), whereas figures on the right each illustrate changes of counterfactual life expectancy (resulting mainly from changes in satisfaction prevalence). Total life expectancy is divided into the number of years that a 60-year-old survivor can expect to live in the very satisfied (life satisfaction of 9 or 10), moderately satisfied (life satisfaction of 7 or 8), and dissatisfied (life satisfaction of 0 to 6) states.
Sources: SOEPv30 (1984-2011) and German Statistical Office 2012a,b, own calculations.

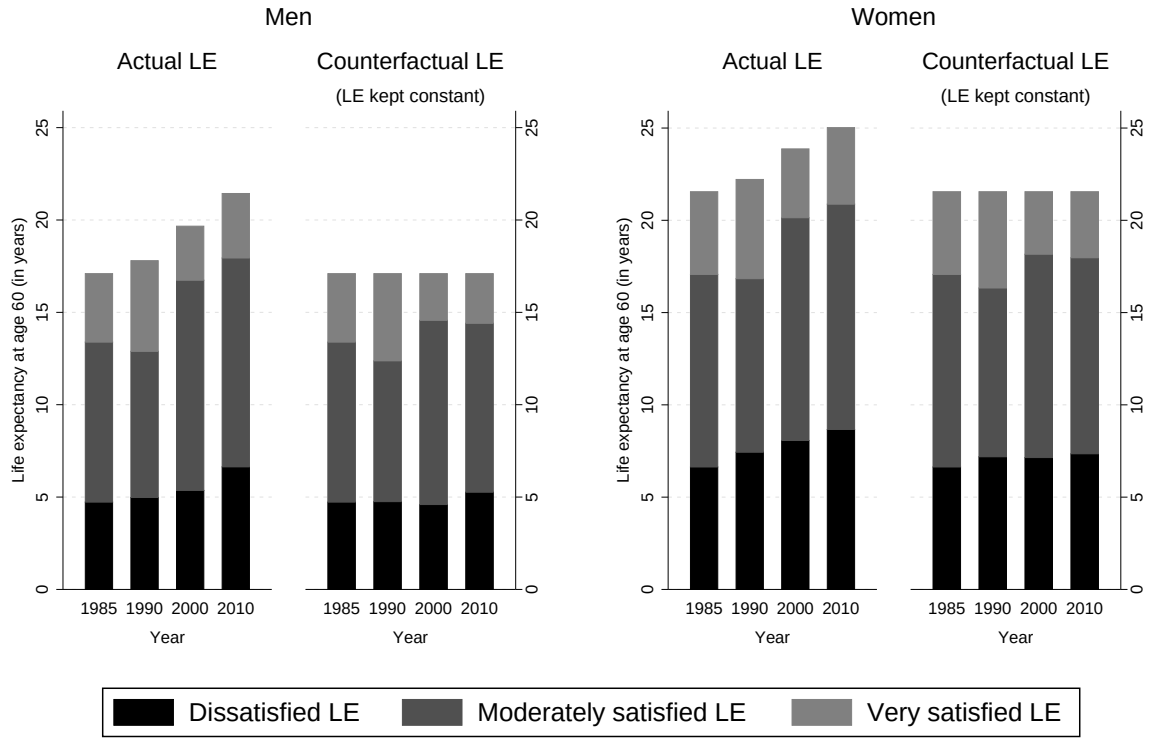

Figure 2B.5. Role of Time-in-Panel Effects: Actual and Counterfactual Satisfied Life Expectancy at Age 60 by Gender and Year

Notes: LE = life expectancy. These figures show the evolution of total and satisfied life expectancy at age 60 in West Germany by gender, after adjusting satisfaction prevalence rates for time-in-panel effects in the estimations. To adjust for time-in-panel effects, I linearly control for time-in-panel duration. Figures on the left each illustrate changes of actual life expectancy (resulting from changes in both satisfaction prevalence *and* mortality), whereas figures on the right each illustrate changes of counterfactual life expectancy (resulting mainly from changes in satisfaction prevalence). Total life expectancy is divided into the number of years that a 60-year-old survivor can expect to live in the very satisfied (life satisfaction of 9 or 10), moderately satisfied (life satisfaction of 7 or 8), and dissatisfied (life satisfaction of 0 to 6) states.
Sources: SOEPv30 (1984-2011) and German Statistical Office 2012a,b, own calculations.

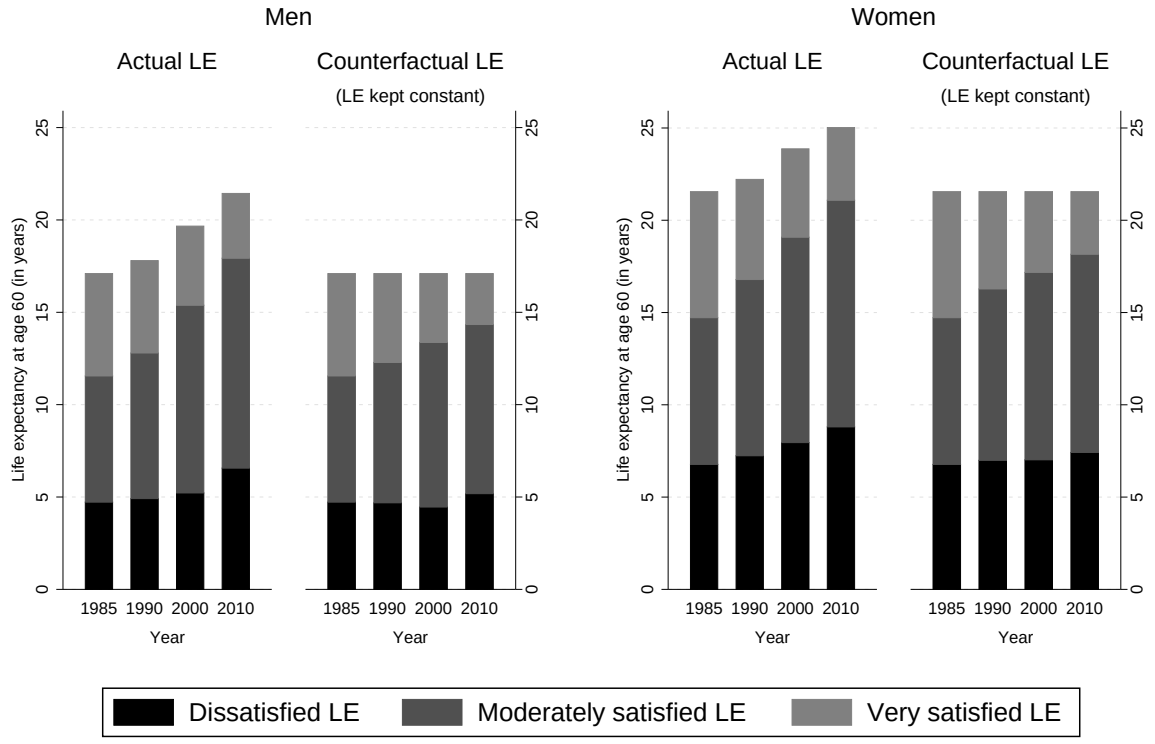

Figure 2B.6. Unweighted Estimates: Actual and Counterfactual Satisfied Life Expectancy at Age 60 by Gender and Year

Notes: LE = life expectancy. These figures show the evolution of total and satisfied life expectancy at age 60 in West Germany by gender, if estimating satisfaction prevalence rates without cross-sectional survey weights. Figures on the left each illustrate changes of actual life expectancy (resulting from changes in both satisfaction prevalence *and* mortality), whereas figures on the right each illustrate changes of counterfactual life expectancy (resulting mainly from changes in satisfaction prevalence). Total life expectancy is divided into the number of years that a 60-year-old survivor can expect to live in the very satisfied (life satisfaction of 9 or 10), moderately satisfied (life satisfaction of 7 or 8), and dissatisfied (life satisfaction of 0 to 6) states.
Sources: SOEPv30 (1984-2011) and German Statistical Office 2012a,b, own calculations.

2C Standard Error of Satisfied Life Expectancy at Age 60

In this appendix, I derive the standard error of satisfied life expectancy at age 60, which is used to test for significant differences across time. Using Chiang's (1984) result, I first rewrite the formula of satisfied life expectancy in terms of survival probabilities, p_{x,n_x}. Then, I use the delta method to obtain the standard errors for this non-linear function of random variables.

Chiang (1984) showed that the person-years lived in the age interval $[x, x+n_x)$, L_{x,n_x}, are a linear function of the cumulative survival probability up to age x, which itself is a product of the probabilities of surviving from each starting age x to age $x+n_x$, p_{x,n_x}:

$$L_{x,n_x} = n_x l_{x+n_x} + n_x a_x (l_x - l_{x+n_x}) \tag{2C.1}$$

$$l_x = l_0 \prod_{i \in B_x} p_{i,n_i}, \tag{2C.2}$$

where $B_x = \{i \in B : x > i\}$ and a_x the average proportion lived by people who die in the age interval $[x, x+n_x)$. Inserting equations (2C.1) and (2C.2) in the formula of satisfied life expectancy at age 60, I obtain

$$e^s_{60} = \frac{\sum_{x \in A_{60}} h^s_{x,n_x} [n_x \prod_{i \in B_{x+n_x}} p_{i,n_i} + n_x a_x (l_0 \prod_{i \in B_x} p_{i,n_i} - l_0 \prod_{i \in B_{x+n_x}} p_{i,n_i})]}{l_0 \prod_{i \in B_{60}} p_{i,n_i}} \tag{2C.3}$$

That is, satisfied life expectancy at age 60 is a nonlinear function of the age-specific survival probabilities, p_{x,n_x}, and the satisfaction prevalence rates, h^s_{x,n_x}, all of which are random variables.

Next, the delta method is applied to this non-linear function of random variables. As the age-specific satisfaction prevalence rates are estimated based on a different data source than the age-specific survival probabilities, they can be considered independent of the age-specific survival probabilities and the covariance terms between these variables can be ignored (Mathers 1991). Moreover, given that the survival probabilities for two non-overlapping age intervals are estimated based on two distinct groups of people, the estimated survival probabilities are uncorrelated across age intervals (Chiang 1960). This argument also holds for German period life tables, which rely on repeated cross-sectional data and pool the data of three years to obtain the life table estimates for a given year.

A similar argument would apply to the age-specific sample fractions of satisfied survey respondents, if pooled data of repeated cross-sections were used. In my case, however, this argument does not apply because I use longitudinal data *and* compute three-year averages. Taking into account that I use five-year age intervals, the sample fractions of satisfied SOEP respondents at a given point in time are correlated across two adjacent age intervals, while they continue to be uncorrelated across the other non-overlapping age intervals. Thus, the delta method yields the following variance of satisfied life expectancy at age 60:

$$\begin{aligned} Var\left(e^s_{60}\right) = &\sum_{x\in A_{60}} \left(\frac{\partial e^s_{60}}{\partial p_{x,n_x}}\right)^2 Var\left(p_{x,n_x}\right) + \sum_{x\in A_{60}} \left(\frac{\partial e^s_{60}}{\partial h^s_{x,n_x}}\right)^2 Var\left(h^s_{x,n_x}\right) \\ &+ 2\sum_{x\in A_{60}\setminus w} \frac{\partial e^s_{60}}{\partial h^s_{x,n_x}} \frac{\partial e^s_{60}}{\partial h^s_{x+n_x,n_{x+n_x}}} Cov\left(h^s_{x,n_x}, h^s_{x+n_x,n_{x+n_x}}\right), \end{aligned} \tag{2C.4}$$

where w is the starting age for the oldest age interval. The first term describes the variation in survival (or mortality), while the second and third term describe the variation in satisfaction prevalence.

According to Newman (1988), the variation resulting from mortality rates will be negligible if the sample size of the survey population relative to the sample size of the population on which the mortality data are based is small. Therefore, I ignore the first term in equation (2C.4).[39] After explicitly writing down the derivatives, the standard error of satisfied life expectancy at age 60 is then given by

$$\begin{aligned} se\left(e^s_{60}\right) = \Bigg(&\frac{\sum_{x\in A_{60}} \left(\frac{L_{x,n_x}}{l_{60}}\right)^2 Var\left(h^s_{x,n_x}\right)}{n} \\ &+ \frac{2\sum_{x\in A_{60}\setminus w} \frac{L_{x,n_x}}{l_{60}} \frac{L_{x+n_x,n_{x+n_x}}}{l_{60}} Cov\left(h^s_{x,n_x}, h^s_{x+n_x,n_{x+n_x}}\right)}{n} \Bigg)^{\frac{1}{2}}, \end{aligned} \tag{2C.5}$$

where n is the sample size of the survey population which is used to estimate the satisfaction prevalence rates, h^s_{x,n_x}. The estimator is obtained by using the information from period life tables and replacing the population variances and covariances in the final equation by their sample counterparts. Unlike in studies that use repeated cross-sectional

[39] In practice, this is a suitable approach, which under the stated conditions produces almost the same standard errors as if the first term was not ignored (Jagger et al. 2014).

data, I compute clustered variances and covariances for the satisfaction prevalence rates in a given year to account for serial correlation across observations of the same respondent within age intervals and across two adjacent age intervals.

2D Mechanisms: Supplementary Tables and Figures

Figure 2D.1. Role of Health: Year Dummy Coefficient Estimates

Notes: These figures illustrate the change of the year dummy coefficient estimates after adding health satisfaction or indicators for the disability status to the adjusted individual-level life satisfaction regressions. Coefficient estimates for the adjusted baseline were obtained from weighted linear regressions of life satisfaction on a set of year dummies, controls for background and survey characteristics, five-year cohort dummies, and time-in-panel duration (cf. final specification in Figure 2A.2 and Table 2A.1 in Appendix 2A). Separate regressions were estimated by time to death. Coefficient estimates are relative to the year 1985, which is the base year in all regressions. Regression results are reported in Table 2D.1 in Appendix 2D.
Sources: SOEPv30 (1984-2013) and GBE (2016), own calculations.

Table 2D.1. Role of Health: Life Satisfaction Regressions by Time to Death

	1 Year			2 or 3 Years			4 or 5 Years		
	Adj. Baseline	Add Health I	Add Health II	Adj. Baseline	Add Health I	Add Health II	Adj. Baseline	Add Health I	Add Health II
Constant	5.509***	4.275***	6.047***	5.426***	3.092***	4.993***	4.550***	2.220**	3.905***
	(0.836)	(0.568)	(0.926)	(0.712)	(0.797)	(0.683)	(0.908)	(0.904)	(0.631)
1986	-0.279	-0.185	-0.356	-0.103	-0.117	-0.198	0.024	-0.027	0.147
	(0.172)	(0.136)	(0.247)	(0.106)	(0.090)	(0.156)	(0.092)	(0.073)	(0.161)
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1997	-0.490	-0.552*	-0.518	-0.260	-0.203	-0.225	0.081	-0.041	0.244
	(0.356)	(0.320)	(0.423)	(0.344)	(0.259)	(0.318)	(0.260)	(0.207)	(0.289)
1998	-0.637*	-0.623*	-0.661	-0.409	-0.278	-0.445	0.088	-0.075	0.243
	(0.373)	(0.319)	(0.433)	(0.344)	(0.261)	(0.312)	(0.253)	(0.199)	(0.284)
1999	-0.738**	-0.763**	-0.732*	-0.423	-0.315	-0.530*	0.057	-0.060	0.252
	(0.363)	(0.306)	(0.425)	(0.330)	(0.252)	(0.300)	(0.244)	(0.195)	(0.276)
2000	-0.780**	-0.819***	-0.780*	-0.264	-0.251	-0.400	0.011	-0.053	0.223
	(0.356)	(0.293)	(0.415)	(0.320)	(0.241)	(0.293)	(0.239)	(0.193)	(0.273)
2001	-0.636*	-0.770***	-0.687*	-0.261	-0.317	-0.353	-0.092	-0.114	0.136
	(0.346)	(0.292)	(0.408)	(0.317)	(0.234)	(0.289)	(0.236)	(0.189)	(0.274)
2002	-0.585*	-0.795***	-0.640	-0.313	-0.308	-0.357	-0.144	-0.128	0.097
	(0.354)	(0.299)	(0.416)	(0.319)	(0.236)	(0.292)	(0.244)	(0.195)	(0.282)
2003	-0.732**	-0.775***	-0.757*	-0.559*	-0.404*	-0.549*	-0.246	-0.074	-0.004
	(0.357)	(0.298)	(0.419)	(0.322)	(0.242)	(0.296)	(0.256)	(0.207)	(0.292)
2004	-0.438	-0.536*	-0.505	-0.728**	-0.451*	-0.719**	-0.201	0.049	0.080
	(0.357)	(0.294)	(0.420)	(0.330)	(0.248)	(0.303)	(0.264)	(0.219)	(0.297)
2005	-0.535	-0.592*	-0.552	-0.780**	-0.482*	-0.803***	-0.257	0.062	0.045
	(0.371)	(0.310)	(0.433)	(0.335)	(0.253)	(0.308)	(0.269)	(0.223)	(0.302)
2006	-0.433	-0.549*	-0.514	-0.698**	-0.445*	-0.738**	-0.171	0.057	0.142
	(0.371)	(0.310)	(0.433)	(0.341)	(0.254)	(0.313)	(0.268)	(0.224)	(0.302)
2007	-0.573	-0.686**	-0.622	-0.576*	-0.405	-0.613*	-0.079	0.036	0.203
	(0.402)	(0.337)	(0.461)	(0.345)	(0.255)	(0.321)	(0.271)	(0.228)	(0.304)
2008	-0.232	-0.376	-0.272	-0.510	-0.366	-0.533			
	(0.401)	(0.338)	(0.460)	(0.349)	(0.260)	(0.325)			
2009	-0.024	-0.282	-0.099	-0.405	-0.215	-0.463			
	(0.423)	(0.357)	(0.484)	(0.351)	(0.267)	(0.328)			
2010	-0.002	-0.349	-0.058						
	(0.416)	(0.346)	(0.474)						
2011	-0.357	-0.722**	-0.415						
	(0.425)	(0.344)	(0.481)						
Health sat.		0.560***			0.509***			0.505***	
		(0.024)			(0.016)			(0.018)	
Disability status (ref. group: not disabled)									
Low			0.041			-0.240			0.062
			(0.382)			(0.318)			(0.258)
Medium			-0.557***			-0.441***			-0.564***
			(0.208)			(0.134)			(0.150)
High			-0.812***			-0.930***			-0.819***
			(0.244)			(0.196)			(0.184)
Fully dis.			-1.902***			-1.706***			-1.937***
			(0.218)			(0.184)			(0.202)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cohort eff.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
TiP effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R2	0.06	0.38	0.13	0.05	0.38	0.11	0.05	0.38	0.13
N	5,221	5,221	4,613	10,532	10,532	9,125	9,797	9,797	8,537

Notes: TiP = time-in-panel. Controls include an indicator for males, years of education, need-weighted deflated net household income, indicators for the interview month, and indicators for the state of residence. Cohort effects are controlled for by including five-year cohort dummies. Time-in-panel effects are controlled for linearly. Low disability (disability degree of 1 to 49%), medium disability (disability degree of 50 to 79%), high disability (disability degree of 80 to 99%), and fully disabled (disability degree of 100%). Standard errors that allow for correlation across observations of the same respondent are reported in parentheses. For a graphical representation of all year dummy coefficient estimates, see Figure 2D.1 in Appendix 2D.

* p<0.1 ** p<0.05 *** p<0.01.

Sources: SOEPv30 (1984-2013) and GBE (2016), own calculations.

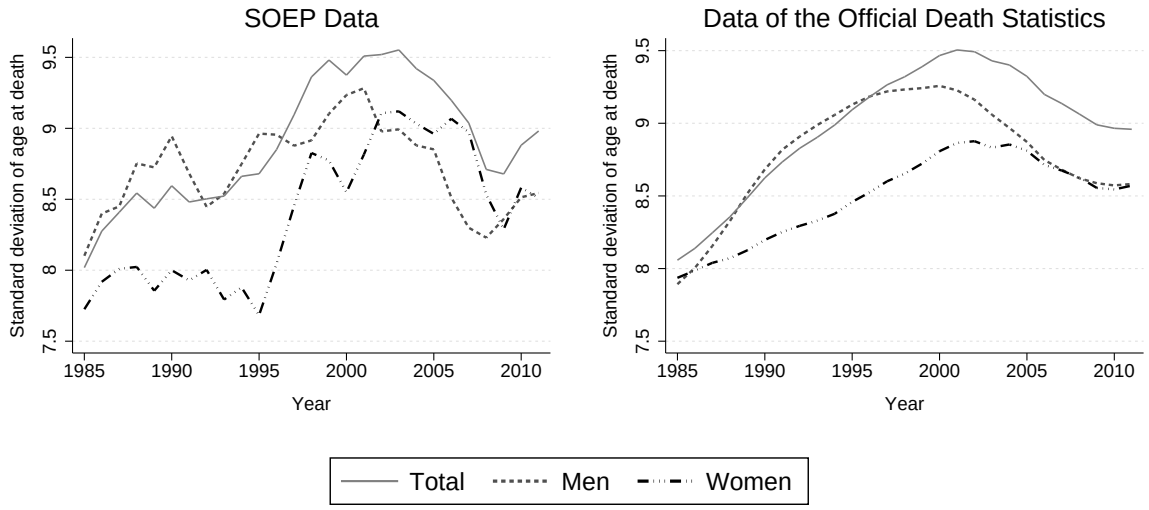

Figure 2D.2. Time Trends of the Mortality Variability for 60-Year-Old Survivors, Total and by Gender

Notes: These figures depict time trends of the standard deviation of the age at death for 60 year-old survivors using (i) SOEP data and (ii) data of the official German death statistics. SOEP estimates are population-weighted and refer to West Germans without migration backgrounds who died at age 60 or older. For the death statistics, estimates refer to all people in Germany who died at age 60 or older. As before, I pool data of three years including and surrounding a particular survey to obtain the estimates in a given year.
Source: SOEPv30 (1984-2013) and German Statistical Office (2016a), own calculations.

Table 2D.2. Role of Social Isolation: Pooled Life Satisfaction Regressions

		Controlled for mutual visits with	
	Adj. Baseline	Family	Friends
Constant	5.061***	4.116***	4.525***
	(0.603)	(0.641)	(0.620)
1995/98	-0.462**	-0.403*	-0.410**
	(0.210)	(0.207)	(0.205)
2003/08	-0.451*	-0.317	-0.312
	(0.238)	(0.238)	(0.231)
Frequency of visits (ref. group: never)			
Less than monthly		0.557	0.512*
		(0.355)	(0.281)
Monthly		1.263***	0.869***
		(0.349)	(0.283)
Weekly		1.246***	1.383***
		(0.333)	(0.266)
Controls	Yes	Yes	Yes
Cohort effects	Yes	Yes	Yes
TiP effects	Yes	Yes	Yes
Adj. R2	0.04	0.07	0.08
N	1,638	1,638	1,638

Notes: TiP = time-in-panel. Pooled regressions for respondents within five years of death. Controls include an indicator for males, years of education, need-weighted deflated net household income, indicators for the interview month, and indicators for the state of residence. Cohort effects are controlled for by including five-year cohort dummies. Time-in-panel effects are controlled for linearly. Standard errors that allow for correlation across observations of the same respondent are reported in parentheses. * p<0.1 ** p<0.05 *** p<0.01.

Sources: SOEPv30 (1984-2013) and GBE (2016), own calculations.

Table 2D.3. Role of Social Isolation: Life Satisfaction Regressions by Time to Death

	1 Year			2 or 3 Years			4 or 5 years		
		Controlled for mutual visits with			Controlled for mutual visits with			Controlled for mutual visits with	
	Baseline	Family	Friends	Baseline	Family	Friends	Baseline	Family	Friends
Constant	5.334***	4.737***	4.560***	5.061***	3.375***	4.536***	5.464***	4.989***	5.172***
	(1.397)	(1.500)	(1.648)	(0.867)	(0.891)	(0.863)	(0.974)	(1.067)	(0.995)
1995/98	-0.130	-0.147	-0.055	-0.612**	-0.550*	-0.573*	-0.445	-0.400	-0.386
	(0.529)	(0.530)	(0.520)	(0.296)	(0.288)	(0.293)	(0.341)	(0.336)	(0.327)
2003/08	-0.795	-0.742	-0.501	-0.441	-0.250	-0.319	-0.248	-0.186	-0.155
	(0.726)	(0.749)	(0.716)	(0.322)	(0.319)	(0.321)	(0.355)	(0.351)	(0.347)
Frequency of visits (ref. group: never)									
Less than monthly		-0.189	0.549		1.395***	0.401		0.346	0.293
		(0.784)	(0.658)		(0.452)	(0.356)		(0.552)	(0.440)
Monthly		0.708	0.879		2.110***	0.817**		0.751	0.596
		(0.745)	(0.689)		(0.436)	(0.381)		(0.530)	(0.437)
Weekly		0.939	1.937***		2.011***	1.020***		0.847*	1.232***
		(0.766)	(0.598)		(0.411)	(0.355)		(0.502)	(0.418)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cohort effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
TiP effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R2	0.01	0.04	0.08	0.04	0.10	0.06	0.04	0.04	0.07
N	317	317	317	675	675	675	646	646	646

Notes: TiP = time-in-panel. Controls include an indicator for males, years of education, need-weighted deflated net household income, indicators for the interview month, and indicators for the state of residence. Cohort effects are controlled for by including five-year cohort dummies. Time-in-panel effects are controlled for linearly. Standard errors that allow for correlation across observations of the same respondent are reported in parentheses. * p<0.1 ** p<0.05 *** p<0.01.

Sources: SOEPv30 (1984-2013) and GBE (2016), own calculations.

Chapter 3

The Effect of All-Day Primary School Programs on Maternal Labor Supply

This chapter is a revised version of Working Paper No. 213 published in the *Working Paper Series* of the Department of Economics, University of Zurich.

Abstract: This study analyzes the effect of all-day primary school programs (ADSP) on maternal labor supply. To account for selectivity of schools with ADSP and selection into ADSPs, I estimate bivariate probit models. To identify these models, I exploit variation in the allocation of investments that were used to set up ADSPs at primary schools across time and counties. This variation results from the public investment program "Future Education and Care" which was introduced by the German federal government in 2003. My results indicate for mothers with primary school-aged children in Germany a significantly positive effect of ADSPs on labor supply at the extensive margin. On average, mothers who make use of ADSPs are 25.2 percentage points more likely to be employed than mothers who do not make use of these programs. This large effect is concentrated among mothers who hold at most a vocational degree and it is robust to alternative specifications. On the contrary, there is no evidence for an impact of ADSPs on maternal labor supply at the intensive margin (full-time vs. part-time).

JEL classification: J13, J21, J22
Keywords: All-day school programs, after-school care, maternal labor supply

Acknowledgements: I wish to thank Uschi Backes-Gellner, Pietro Biroli, Richard Blundell, Stefan Boes, Steven Haider, Kristina Hess, Christian Kascha, Johannes Kunz, Steffen Künn, Edward Lazear, Harald Mayr, Florian Schaffner, Rainer Winkelmann, Susanna Zitz, seminar participants at Grindelwald, Lech am Arlberg, Lucerne, Zurich, the "Zurich Workshop in Economics", and the Workshop on "Empirical Analysis of Consumption and Labour Supply Behaviour" at the Study Center Gerzensee, and participants at the 2016 annual meeting of the German Economic Association in Augsburg and the 28th EALE conference in Ghent for helpful comments and suggestions, and Luca Tonizzo for excellent research assistance.

3.1 Introduction

In the last three decades, many OECD countries have introduced childcare programs to enable mothers to better reconcile work and family life. These programs have reduced the costs associated with childcare by either expanding overall childcare supply or subsidizing the utilization of childcare. As the reduction of childcare costs increases the net benefits of employment (Becker 1991, Ribar 1992), childcare programs are expected to increase maternal labor supply. Poor childcare quality or high social costs of using childcare (e.g., loss in social esteem, "Rabenmutter") may, however, suppress maternal acceptance of childcare (Blau and Robins 1988, Ribar 1992). A priori, it is thus unclear whether or not maternal labor supply indeed increases in response to childcare programs.

For this reason, many studies have analyzed the impact of childcare programs on maternal labor supply (e.g., Gelbach 2002, Berlinski and Galiani 2007, Baker et al. 2008, Cascio 2009, Fitzpatrick 2010, Havnes and Mogstad 2011, and Bauernschuster and Schlotter 2015). These studies have primarily focused on the effects of care for preschool-aged children, notwithstanding that childcare also plays an important role for young school-aged children, in particular, in countries with a half-day school system. Figure 3.1 illustrates this for Germany. According to the German time use survey 1991/92, the time that mothers spent with children decreases with the age of the youngest child in the household (figure a). On the contrary, maternal employment shares and the share of full-time employed mothers on all mothers increase with the age of the youngest child in the household (figure b). Taken together, these two figures suggest that, unlike for older school-aged children, care for young school-aged children is still strongly linked to maternal labor supply.

To the best of my knowledge, there exist only three studies that evaluate the effect of childcare provision for (young) school-aged children on maternal labor supply. Felfe et al. (2016) exploit cantonal variation in the regulations of after-school care provision in Switzerland to show that mothers with 4- to 12-year-old children positively respond to an increased after-school care coverage at the intensive margin. Berthelon et al. (2015) and Contreras and Sepúlveda (2017) study the effect of an extension of school schedules from half to full day on maternal labor supply in Chile. Both studies provide evidence for a substantial increase in maternal labor force participation, but neither of them finds

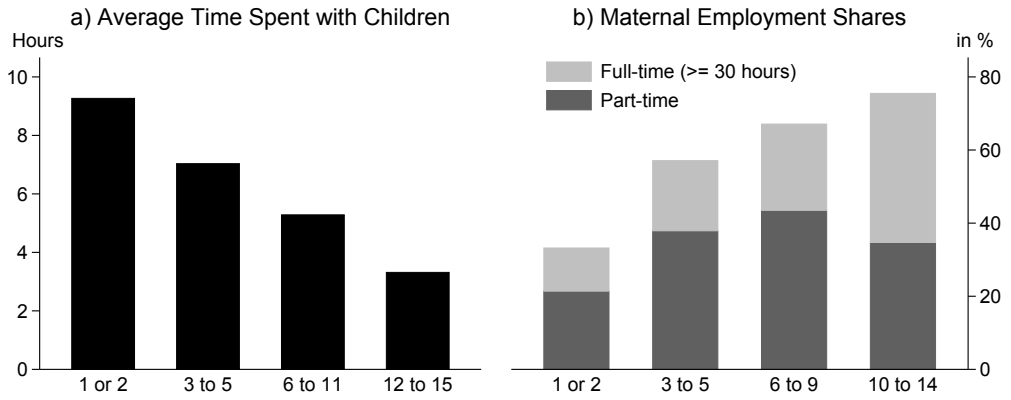

FIGURE 3.1. MATERNAL CARE AND EMPLOYMENT BY AGE OF THE YOUNGEST CHILD

Sources: a) BMFSFJ (1996): based on data of the German time use survey 1991/92. b) Kreyenfeld and Geisler (2006): based on data of the Microcensus 2002. Own representation.

a positive effect of the prolonged school day on maternal labor supply at the intensive margin.

In this study, I analyze whether voluntary all-day school programs (ADSPs) for primary school-aged children increase the labor supply of mothers with primary school-aged children in Germany. In 2003, the German federal government launched the public investment program "Future Education and Care"[1] (IZBB), which has led to a sharp increase of ADSPs in Germany. Between 2003 and 2009, the share of primary schools with ADSPs more than tripled, reaching 41.8% in 2009 (KMK 2014). Yet, the amount invested into the expansion of ADSPs differed across counties and over time. I exploit this quasi-experimental setting to identify causal effects in the presence of selectivity of schools with ADSP and selection into ADSPs.[2] I jointly model maternal labor supply and ADSP use and estimate two bivariate probit models – one model for the extensive margin (employment vs. non-employment) and one model for the intensive margin (part-time vs. full-time employment conditional on maternal employment). In order to identify these models, I use exogenous variation in the cumulative amount of IZBB investments per primary school across counties and over time, which resulted from gradual implementation of the IZBB program. The analyses are based on data from the German Socio-Economic Panel (SOEP).

[1] In German: "Zukunft Bildung und Betreuung".

[2] As the supply of ADSPs was very low at the beginning of the IZBB investment period, schools gave priority to particular types of families if capacity constraints of a program were reached. Moreover, as parents were able to decide according to their preferences upon the use of ADSPs for their primary school-aged children, parents who selected into these programs likely differ from parents who did not. For a more detailed discussion, see section 3.4.1.

As a primary result, I find that ADSPs at the primary education level substantially increase maternal labor supply at the extensive margin, while they have no effect on maternal labor supply at the intensive margin. When compared to mothers with primary school-aged children who do not make use of ADSPs, mothers with primary school-aged children who make use of ADSPs have a 25.2 percentage points higher probability of being employed. This large effect is concentrated among mothers with at most a vocational degree, and it is robust to a number of alternative specifications. Two-stage least-squares (2SLS) estimates lack precision, but they point in the same direction as the bivariate probit estimates.

As a secondary result, I find that selection (on unobservables) into ADSPs is negative. This finding is compatible with some anecdotal evidence on the acceptance of ADSPs in Germany. As ADSPs at the primary education level were made available to parents on a voluntary basis, no extra lessons were offered in the afternoon. Instead, schools offered a broad variety of social and cultural activities at essentially zero cost, which rendered these programs particularly attractive to parents with low socio-economic status and tighter budget constraints. This anecdotal evidence is supported by Börner et al. (2010) who find that better educated parents and parents with high socio-economic status often critically oppose ADSPs, while parents with low socio-economic status and/or migration backgrounds often believe that their child benefits from the attendance of ADSPs.

This study contributes to the existing literature in at least three important ways. First, this study is the first analyzing whether ADSPs, which are made available to parents on a voluntary basis, increase the labor supply of mothers with primary school-aged children.[3] This research question has not been addressed yet as the attendance of ADSPs was mandated in Chile. This study corroborates the Chilean findings and demonstrates that they probably continue to hold if parents were allowed to opt out of these programs. Second, although the activities offered by ADSPs in Germany resemble those of after-

[3]Note that Beblo et al. (2005) evaluate the potential impact of the expansion of ADSPs at the primary education level on maternal labor supply in an ex-ante simulation study. Initial descriptive evidence on the impact of ADSPs on maternal labor supply is, for example, provided in Tobsch (2013) and Rainer et al. (2013). The same quasi-experimental setting is used in three parallel studies (Shure 2016, Gambaro et al. 2016, and Dehos and Paul 2017), which use different samples (in order of studies: focus on four German states only, mothers whose children enter primary school, youngest primary-school aged child in household) and estimation methods (difference-in-difference, matching, and two-sample 2SLS), and identify different treatment effects (intention to treat effect, average treatment effect on the treated, local average treatment effect), when compared to this study. Despite these differences, our results show partial (Gambaro et al. (2016) for the extensive margin and Dehos and Paul (2017) for the intensive margin) to full agreement (Shure 2016). For an earlier version of this paper, see Nemitz (2015).

school care, maternal labor supply responses to ADSPs likely differ from those uncovered for after-school care in Felfe et al. (2016). This is because, unlike after-school care programs, which usually last until 6 p.m., ADSPs generally finish between 3 p.m. and 4 p.m. Therefore, despite increasing the time spent in schools, ADSPs often remain incompatible with full-time working schedules. Third, by analyzing differential labor supply responses of mothers by education, this study finds that only lower educated mothers respond to ADSPs by resuming part-time employment. This finding is novel in that it suggests that childcare provision may increase rather than decrease the gender gap in working hours and wages if operating hours of childcare remain incompatible with full-time working schedules.[4]

The remainder of this paper is structured as follows. Section 3.2 provides some information on the institutional framework in Germany and the IZBB program. Section 3.3 describes the data and sample selection. Section 3.4 presents the empirical strategy and discusses the main underlying identifying assumption. Section 3.5 presents the results and a battery of robustness checks. Section 3.6 concludes.

3.2 Institutional Background

In this section, I provide some general information on the institutional background in Germany. In particular, I give some details on the primary education system, care for primary school-aged children, and maternal employment prior to the expansion of ADSPs in Germany. I proceed by presenting the IZBB program, which has led to the expansion of ADSPs. Finally, I briefly discuss the structure of ADSPs in Germany.

3.2.1 Institutional Setting Before the IZBB Program

In Germany compulsory primary education starts when children turn six years old and, in general, it lasts for four years. Therefore, the majority of children completes primary school by the age of ten. Prior to the IZBB program, primary schools were mainly organized as half-day schools. Yet, the exact time spent in primary schools varied across states and institutions. While several states introduced schedules that guaranteed a supervision of children until 1 p.m. on all days of the week, there were also many schools that offered

[4]In Germany, wages for part-time employed women are, on average, 11% lower than wages for full-time employed women, even after controlling for educational differences (Wolf, 2002, 2010).

flexible schedules, which only started after 9 a.m. or finished already at 11 a.m. on some days of the week. This short and irregular time spent in primary schools was strongly incompatible with regular working schedules in Germany. Even part-time working schedules often clashed with school schedules, since lunch was not served in schools. Therefore, and because of the low supply of after-school center-based care[5], mothers with primary school-aged children faced substantial difficulties in combining work and family life prior to the launch of the IZBB program.

As a consequence, in 2002, the labor force participation rate of women with primary school-aged children was more than ten percentage points lower in Germany than in Scandinavian countries. Yet, with a rate of 73.3% it was comparable to the rates of European countries.[6] Full-time employment rates were substantially lower than in European countries, however. According to the German microcensus, only 35.2% of the women with primary school-aged children were full-time employed in 2002. Thus, in combination with the low supply and use of institutionalized care for young school-aged children these employment statistics suggest that ADSPs for primary school-aged children likely have a relatively large impact on maternal labor supply. This suggestion is reinforced by surveys conducted after the end of the IZBB program. Among non-employed mothers with primary school-aged children, 50% indicated childcare as the main reason for not searching employment (Statistisches Bundesamt 2012). Moreover, one of the main reasons indicated by part-time employed mothers for not taking up full-time employment was the lack of adequate childcare for school-aged children once the school finishes (BMFSFJ 2011).

3.2.2 The IZBB Program and ADSPs in Germany

In order to enhance the compatibility between work and family life, the German federal government launched the public investment program "Future Education and Care" in 2003.[7] The main purpose of this program was the establishment and expansion of ADSPs

[5]Note that there were substantial differences in the supply of after-school center-based care across regions. While in West Germany (excluding the city states) only 6% of 6- to 9-year-old children were offered after-school care in 2002, the corresponding share was more than ten times higher in East Germany due to the socialist heritage (DJI 2005, p. 144). Differences of similar magnitude were observed between rural and urban areas. In West Germany the place-to-child ratio in urban areas was, for example, seven times higher than the ratio in rural areas (2002: place-to-child ratio of 2.1% in rural areas, DJI 2005, p. 145).

[6]This figure and the subsequent one refer to 18- to 45-year-old women whose youngest child was aged six to nine. They were calculated based on microcensus data by Kreyenfeld and Geisler (2006).

[7]Note that this program also had other objectives. In particular, German politicians aimed at achieving more educational and social justice. For a summary on all objectives of the IZBB program, see BMFSFJ (2013).

in Germany in order to create a demand-oriented and area-wide supply of ADSPs in all states (BMBF 2003). The total investment volume of the IZBB program was about four billion euros. However, as federal investments were made conditional upon additional investments by the states, the actual investment volume was considerably higher. Between 2003 and 2009, IZBB investments were allocated to more than 8200 schools. Primary schools were of particular importance: More than half of the total investment volume was allocated to schools in the primary education sector (BMBF 2009). The share of primary school students participating in ADSPs more than quadrupled since 2003. In 2009, 21.5% of all primary school students in Germany attended an ADSP (KMK 2014).

How are ADSPs expected to affect maternal labor supply? Depending on the state in Germany, ADSPs increase the time spent in primary schools by 30 to 100%. Due to the absence of a general definition of ADSPs at the primary education level, there exists substantial variation in the operating hours of ADSPs, however. Although primary schools, which offer an ADSP, are obliged to provide a program that comprises seven hours per day on at least three days per week (KMK 2014), many schools deviate from this regulation and voluntarily provide a more comprehensive program, which covers all regular working days (Monday to Friday) or goes beyond the seven hours per day. Yet, only in some exceptional cases, primary schools offer a program that ends between 4.30 p.m. and 6 p.m. Therefore, ADSPs often remain incompatible with regular full-time working schedules in Germany. Consequently, I expect ADSPs to have no effect on maternal full-time employment probabilities. ADSPs are, however, very likely to have a large impact on maternal employment probabilities, probably through an increase in maternal part-time employment.

What do ADSPs offer and which mothers are they expected to attract? As ADSPs in the primary education sector are made available to parents on a voluntary basis, schools are not allowed to schedule extra lessons in the afternoon. Instead, many schools offer social and cultural activities throughout the additional hours. These activities are often provided in cooperation with external associations (e.g., sports clubs, music schools). Daily homework assistance is a constituent component of many ADSPs. Lunch is always provided in school under an ADSP. The attendance of ADSPs is highly subsidized. Including the subsidy, prices of ADSPs range from €30 to €150 per month (cf. Dohmen

and Himpele 2006).[8] Therefore, ADSPs are particularly attractive to socio-economically disadvantaged parents. Although ADSPs are often used by parents for other reasons than maternal employment, the primary reason for using an ADSP indicated by mothers is employment (BMFSFJ 2011, p. 21).

3.3 Data

I combine four different data sets, one of which is novel and provides detailed information on schools that received federal investments throughout the IZBB investment period. In the following, I briefly describe these data sets, the data preparation process, and the sample selection procedure.

3.3.1 Data Sources

The main data source is the German Socio-Economic Panel. The SOEP is a representative longitudinal study of private households in Germany, which was launched in 1984. Originally, the survey included West German households only. After the German reunification, a representative sample of East German households was added. Interviews are conducted on an annual basis. In 2013, more than 25,000 persons in more than 13,000 households were interviewed. The survey includes questions on demographics, household composition, educational outcomes, and labor market characteristics of the respondent. Information on the attendance of ADSPs is collected since 1995.[9]

In addition, I use information on the exact amount of IZBB investments allocated to each primary school in a given year.[10] This information was collected by the Social Pedagogical Institute of the Technical University of Applied Science in Cologne during the IZBB investment period (SPI NRW 2010). I adjusted the original data in two ways. First,

[8] Prices of ADSPs differ across states, municipalities, and school authorities. They often gradually increase with parental income. Besides the tuition fee, parents are charged a fee for lunch in school. Sometimes, additional fees apply for extracurricular activities.

[9] For further information on the survey content and the sampling structure of the SOEP, see Wagner et al. (2007).

[10] For Thuringia only information on the years of investment and the cumulative amount of IZBB investments allocated to each school between 2003 and 2009 are available. Hence, to obtain year-specific data on the amount of IZBB investments allocated to each school in a given year, I assume that the cumulative amount of funding allocated to a school was distributed equally across those years that were indicated for IZBB investment receipt. I show later that the results are robust to this assumption, i.e. they continue to hold if I exclude mothers from Thuringia. In addition, note that I only use IZBB investments that were allocated to primary schools, i.e. I do not consider IZBB investments that were allocated to schools with primary *and* secondary education tracks, unless it was clearly indicated that these investments were allocated to the primary education track.

I replaced implausible low IZBB investment values by a value of zero.[11] Second, I replaced negative IZBB investment amounts that were reported for some primary schools in years towards the end of the IZBB investment period by a value of zero and subtracted these negative investment amounts from positive investment amounts that were reported for preceding years. As all primary schools with negative IZBB investment amounts received positive IZBB investments in at least one preceding year, a likely explanation for these negative IZBB investment amounts is an excess provision of IZBB investments which resulted in unused resources. After these adjustments, I aggregated the data at the county level to obtain the total amount of IZBB investments allocated to primary schools in a county in a given year.

In order to compute the cumulative amount of IZBB investments allocated to primary schools between 2003 and year t per primary school at the county level, I merged the IZBB investment data with data on the number of primary schools per county. Yearly data on the number of primary schools at the county level were obtained from federal statistics and official statistics of the states. They were combined in a single data set by ReGENESIS (2016).

Finally, I supplemented the data with county level employment data. The female regional unemployment rate and the female regional part-time employment rate were retrieved from the Federal Institute for Research on Building, Urban Affairs and Spatial Development (BBSR Bonn 2015).

3.3.2 Sample Selection

Throughout the analyses, I focus on mothers in Germany who lived in a private household between 2003 and 2009 and had at least one primary school-aged child. For the sake of comparison across states, I only consider mothers whose children were at the age of five to ten when attending primary school.[12] I excluded Bavarian mothers from the sample due to the lack of yearly data on schools that received IZBB investments in Bavaria.[13]

[11] Implausible investment values range from €0.10 to €5.64 for a given school in a given year. This error appears to be systematic as it almost exclusively concerns primary schools in North Rhine-Westphalia. Given that 7.6% of all school-year-observations in the original data showed implausible low IZBB investment values, I conducted a robustness check that excluded counties that were affected by this possibly erroneous reporting. The results are robust to the exclusion of these counties.

[12] In Berlin and Brandenburg primary education lasts six years, i.e. children complete primary education at the age of 12.

[13] Between 2003 and 2009, the share of IZBB investments allocated to primary schools in Bavaria was only about 5.1% (SPI NRW 2010).

To ensure that mothers belonged to the working age population, I removed four mothers who were younger than 15 or older than 64 from the sample. Furthermore, I discarded 132 mothers who were self-employed or in education. Both groups of mothers tended to be more flexible with respect to their working hours and the choice of their working place, thus facing less difficulty in combining work and family life. Finally, I excluded 194 mothers with missing information in one of the dependent or explanatory variables.[14] After the deletion of observations with missing values, the remaining sample encompasses 1,764 mothers with 5,016 mother-year-observations.

For the descriptive analyses two additional samples were drawn. The first sample also focuses on mothers as the main unit of analysis and is employed to analyze the maternal use of ADSPs and maternal employment over time. The second sample focuses on children of these mothers instead and is employed to investigate whether ADSPs have crowded out alternative types of care for primary school-aged children.[15] I applied the same sample restrictions as for the main sample, except that I expanded the time period to also cover some pre- and post-IZBB investment years. The extended time period covers all years between 1997 and 2013.

3.4 Empirical Strategy

In Germany, the impact of ADSPs on maternal labor supply is likely distorted by two types of selection: selection at the school level (selectivity of schools with ADSP) and selection induced by parents (selection into ADSPs). In order to account for both types of selection, I estimate bivariate probit models and exploit exogenous variation in the allocation of IZBB investments to primary schools across counties and over time. This section discusses the selection problem and presents the empirical strategy.

3.4.1 Selectivity of Schools with ADSP and Selection into ADSPs

In Germany, public schools must virtually accept all students who live within their attendance boundaries. As schools with ADSP were, however, relatively rare at the beginning

[14]The descriptive results are very similar if Bavarian mothers, mothers who were self-employed or in education, and mothers with missing information in one of the explanatory variables are included in the sample. The same holds true for the multivariate results if self-employed mothers and mothers who were in education are included in the sample.

[15]As information on the collection of childcare variables was incomplete for some children, I had to exclude some children of mothers that were included in the first sample used for descriptive purposes.

of the IZBB investment period, these schools were allowed to reject students if capacity constraints of a school were reached. In such a case, priority was often given to particular types of families (e.g., single-parent, dual-earner, and socially or economically disadvantaged families). However, selection criteria differed across schools and little is known about the choices made within each school. Selectivity of schools with ADSP could thus lead to spurious correlation between the ADSP use indicator and the error term in a single-equation probit for maternal (full-time) employment. In particular, I would overestimate the true effect of ADSPs on maternal labor supply if there were some unobserved maternal characteristics that were positively correlated with both the use of ADSPs and maternal (full-time) employment.

Apart from selectivity of schools with ADSP, selection into ADSPs could lead to biased estimates in a single-equation probit model for maternal (full-time) employment. Since in Germany parents can decide according to their preferences upon the use of ADSPs[16], it is likely that mothers with more favorable labor market characteristics and stronger work preferences more often opt for the use of an ADSP. As these mothers also face a higher likelihood of being (full-time) employed, part of the ADSP effect could be due to the way mothers select into ADSPs. Yet, in general, negative selection into ADSPs would also be possible if mothers with more unfavorable labor market characteristics and weaker work preferences had stronger preferences for ADSPs. Stronger preferences for ADSPs may, for example, result from the fact that children of these mothers get access to a broad variety of different activities which they might not have had access to otherwise (cf. Börner et al. 2010). In anticipation of a positive effect of ADSPs on maternal labor supply, one would be particularly concerned about positive selection, however.

Given that different types of selection may partial out each other, the presence of selection has to be empirically tested. On top of controlling for a large battery of covariates in the regressions, I test for selection by estimating bivariate probit models – one for the extensive margin (employment vs. non-employment) and one for the intensive margin (full-time vs. part-time employment conditional on maternal employment). Bivariate pro-

[16]Roughly 90% of all primary schools with ADSP make the ADSP available to parents on a voluntary basis (KMK 2014). Therefore, these schools continue to offer the classical half-day school program besides the ADSP. The remaining 10% of primary schools with ADSP oblige parents to use the ADSP. Yet, if parents are not willing to make use of an ADSP, they can register their child at a different school outside of their school district. Similarly, parents' school choices are not limited by school districts if parents are willing to make use of an ADSP, but the school which is closest to the place of living does not provide such a program.

bit models have the advantages of directly accounting for selection on unobservables and explicitly incorporating the binary nature of dependent variables to produce (full-time) employment probabilities that lie within a reasonable range of values. Bivariate probit models rely on stronger assumptions than alternative methods, however. Therefore, I also report 2SLS estimates as a robustness check, although 2SLS estimates can differ from bivariate probit estimates in the presence of selection on unobservables and under treatment effect heterogeneity.

3.4.2 Bivariate Probit Models

In the following, I describe the model that is employed to estimate the effect of ADSPs on maternal labor supply at the extensive margin.[17] Let the indicator variable $E_{icst} = 1$ if mother i in county c and state s is employed in year t, and let $E_{icst} = 0$ otherwise. Moreover, let $ADSP_{icst} = 1$ if mother i in county c and state s makes use of an ADSP in year t, and let $ADSP_{icst} = 0$ otherwise.[18] Then, the maternal choice problem is described by the following system of equations:

$$
\begin{aligned}
E^*_{icst} &= X'_{icst}\beta + \alpha ADSP_{icst} + \lambda_s + \omega_t + \varepsilon_{icst} && (3.1)\\
ADSP^*_{icst} &= Z'_{icst}\gamma + \lambda_s + \omega_t + u_{icst} && (3.2)\\
E_{icst} &= \mathbf{1}\{E^*_{icst} > 0\} && (3.3)\\
ADSP_{icst} &= \mathbf{1}\{ADSP^*_{icst} > 0\} && (3.4)\\
\begin{pmatrix} \varepsilon_{icst} \\ u_{icst} \end{pmatrix} &\sim BIVN(0, 0, 1, 1, \rho), && (3.5)
\end{aligned}
$$

where E^*_{icst} and $ADSP^*_{icst}$ denote the latent net benefits that mother i in county c and state s receives from engaging in employment and making use of an ADSP in year t, X_{icst}

[17] To obtain the model for the intensive margin, the variable E only needs to be replaced by its counterpart for the intensive margin, FTE. FTE is equal to one if the mother is full-time employed and equal to zero if the mother is part-time employed or marginally employed. As the variable FTE is available for employed mothers only, estimated effects are conditional on maternal employment.

[18] Since in the SOEP information on the use of ADSPs was collected at the child level, I had to aggregate this information at the maternal level for mothers with multiple primary school-aged children. In the main specification, I set the indicator for the use of ADSPs at the maternal level equal to one if at least one primary school-aged child of a mother made use of an ADSP and zero else, but I show later that the results are robust to alternative specifications. For further information on these specifications and the operationalization of the ADSP indicator, see Appendix 3A.

and Z_{icst} are vectors of maternal characteristics[19], λ_s and ω_t represent state and year fixed effects, and ε_{icst} and u_{icst} are random error terms, which are assumed to be bivariate normally distributed. Equations (3.3) and (3.4) state that a mother will only engage in employment if the expected net benefit of being employed is positive and that she will only make use of an ADSP if the expected net benefit of using this program exceeds zero. A detailed description of the dependent variables and all covariates which are used in the analysis is provided in Table 3B.1 in Appendix 3B.

Assuming independence across time and mothers, the log likelihood function for the bivariate probit model is given by

$$\ln L = \sum_{i=1}^{n} \sum_{t=1}^{T_i} \ln \Phi_2(w_{E,icst}, w_{ADSP,icst}, \rho^*), \tag{3.6}$$

where Φ_2 is the cumulative density function of the standard bivariate normal distribution, and

$$\begin{aligned} w_{E,icst} &= (2E_{icst} - 1)(X'_{icst}\beta + \alpha ADSP_{icst} + \lambda_s + \omega_t), \\ w_{ADSP,icst} &= (2ADSP_{icst} - 1)(Z'_{icst}\gamma + \lambda_s + \omega_t), \\ \rho^* &= (2E_{icst} - 1)(2ADSP_{icst} - 1)\rho. \end{aligned}$$

However, the assumption of no correlation over time is overly restrictive. Wooldridge (2010, Ch. 13.8) has shown for the random effects probit model that, under the assumption of normally distributed random effects, a pooled probit model, though misspecified, recovers consistent average partial effects. Therefore, he suggests estimating pooled probit models instead of proper random effects models. Cluster-robust standard errors are required to allow for correct inference. I adopt this strategy and apply it to the bivariate probit model.

The main effect of interest, the average treatment effect (ATE), is computed as the average difference between the probability that a mother would be employed if she made

[19]I control for the following maternal characteristics: age, migration background, marital status, education, experience, experience squared, the number of children in the household, the presence of young children in the household, deflated need-weighted monthly non-wife net household income, and community size. Moreover, I control for the female county-level unemployment rate (extensive margin model), and the female county-level part-time employment rate (intensive margin model).

use of an ADSP and the probability that she would be employed if she did not make use of such a program. Hence, the estimator of the ATE is given by

$$\widehat{ATE} = \frac{1}{\sum_{i=1}^{n} T_i} \sum_{i=1}^{n} \sum_{t=1}^{T_i} [\Phi(X'_{icst}\hat{\beta} + \hat{\alpha} + \hat{\lambda}_s + \hat{\omega}_t) - \Phi(X'_{icst}\hat{\beta} + \hat{\lambda}_s + \hat{\omega}_t)], \tag{3.7}$$

where Φ denotes the cumulative distribution function of the standard normal distribution. I report cluster-robust standard errors for the ATE, which were derived based on the delta method.

Under independence of the structural errors, ε_{icst} and u_{icst}, the bivariate probit model would simplify to two univariate probit models. Hence, an insignificant estimate of the correlation coefficient ρ suggests that a single-equation probit for maternal (full-time) employment uncovers the ATE. I find evidence for substantial selection on unobservables, however. Therefore, bivariate probit models are the preferred estimation strategy.[20]

3.4.3 Identification

In order to achieve identification in the absence of bivariate normally distributed errors, I require at least one variable in Z_{icst} that is excluded from X_{icst}. I use the cumulative amount of IZBB investments allocated to primary schools between 2003 and year t per primary school at the county level as instrument. Substantial variation in cumulative IZBB investments across counties and over time resulted from gradual implementation of the IZBB program. Figure 3.2 illustrates some of this variation. In 2006, cumulative IZBB investments allocated to primary schools substantially differed among the 306 counties (net of Bavarian counties). Within states, differences among counties were sizable, although these differences were often smaller than differences among counties across states.

The main underlying economic idea of this instrument is as follows: As IZBB investments for primary schools were mainly used to set up new ADSPs, IZBB investments improved the access to ADSPs and, thus, reduced the costs associated with ADSP use. As the cost of childcare affect its demand (e.g., Blau and Robins 1988), an increase in the amount of IZBB investments allocated to primary schools in the county of residence should raise the maternal probability of using an ADSP for primary school-aged children.

[20]For the intensive margin model, the estimate of ρ turns out to be significant in some but not all specifications. Nevertheless, I estimate a bivariate probit model to ensure that the results for the intensive margin also hold in the presence of selection on unobservables.

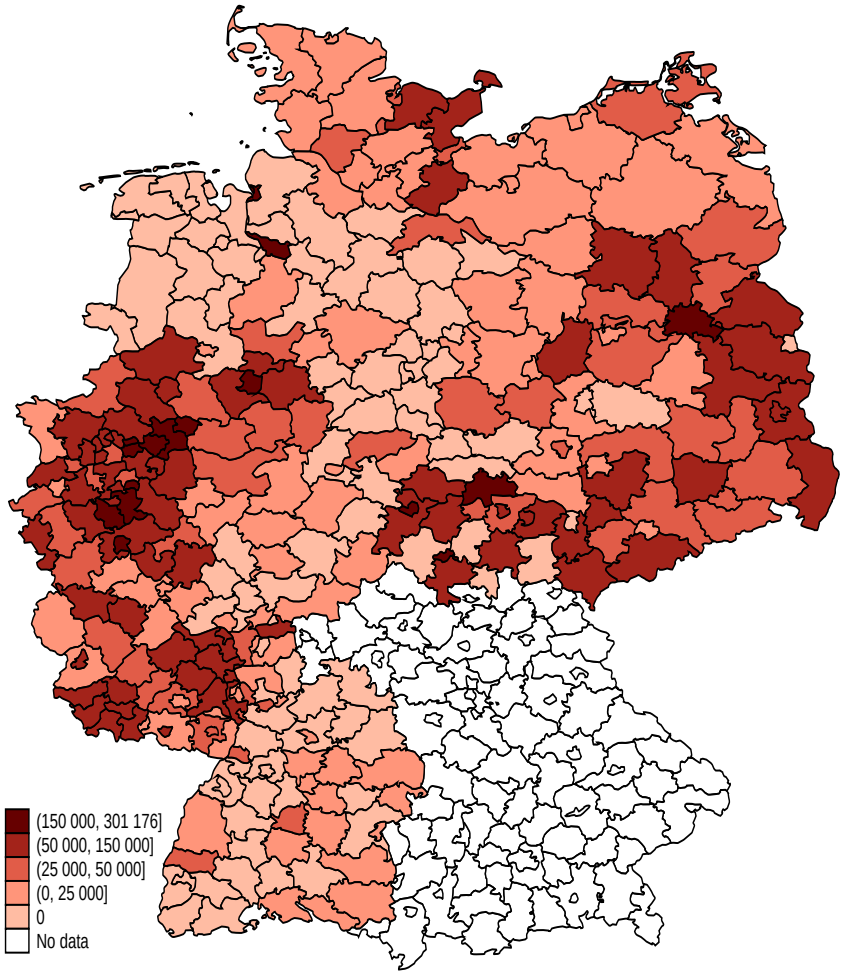

Figure 3.2. Amount of Cumulative IZBB Investments per Primary School in Euros (County-Level), Germany 2006

Sources: SPI NRW (2010) & ReGENESIS (2016), own representation.

The magnitude of effects, however, depends on the overall availability of ADSPs in a county (Cascio et al. 2015), i.e. it tends to be larger if the share of primary schools with ADSP on all primary schools is lower. Therefore, I use the cumulative amount of IZBB investments between 2003 and year t rather than the year-specific amount of IZBB investments.[21] Moreover, as counties differ with regard to their size and number of potential recipients, I divide the cumulative amount of IZBB investments by the number of primary schools in a county.

The main underlying assumption, which has to hold in order to ensure the validity of this instrument, is the exclusionary restriction. This restriction requires that cumu-

[21] I show later that this specification is strongly supported by the data, since lagged IZBB investments predict current maternal use of ADSPs for primary school-aged children, even conditional on current IZBB investments. One likely explanation for the relevance of lagged investments is that IZBB investments were mainly used for constructional purposes. Personal and operating costs of ADSPs were born by the states and communities, using alternative financial resources.

lative IZBB investments per primary school at the county level are unrelated with the error term in the maternal (full-time) employment equation, i.e. after conditioning on covariates in the regressions, cumulative IZBB investments per primary school should only affect maternal (full-time) employment via the use of ADSPs. Thus, a major threat to the empirical strategy would be if mothers were able to influence the allocation of IZBB investments or if school officials rationally sought funding where they anticipated a high demand (e.g., in counties with high maternal employment). The exclusionary restriction is fundamentally untestable. In the following, I argue, however, that this assumption is very likely to hold in the German context after conditioning on state fixed effects. In addition, I provide some suggestive empirical evidence for the validity of this assumption.

As in Germany education policy is determined at the state level, the allocation of IZBB investments to primary schools across states was non-random. This was for two reasons: First, the total amount of federal investments provided to states in a given year differed across states, since it was proportional to the total number of school students in each state in 2001/02. Second, the share of federal IZBB investments allocated to primary schools varied across states due to different targets of the states, which were declared at the beginning of the IZBB investment period (cf. BKJ 2006, Bertelsmann Stiftung 2012).[22] While the former is rather unlikely to pose a threat to the empirical strategy, the latter could be problematic if the targets of the states were systematically linked to the (full-time) employment probabilities of mothers with primary school-aged children. To address this issue, I condition on state fixed effects in all empirical models and only exploit county-level variation of IZBB investments within states.[23]

Within states, the allocation of IZBB investments was likely random because the amount of IZBB investments allocated to a primary school depended on the size and the number of other applicant schools in a state. Moreover, the possibility to apply for IZBB investments was strongly linked to the financial resources of the school operator and cooperation opportunities with external associations (e.g., sports clubs, music schools) as IZBB investments were only granted conditional upon sufficient additional financial and

[22]Some states mainly focused on the expansion of ADSPs in the primary education sector (e.g., Berlin and North Rhine-Westphalia), while other states did so for the secondary education sector (e.g., Bavaria, Baden-Württemberg, and Lower-Saxony). Beyond that, there was a third group of states which focused on the establishment of ADSPs at schools that were located in areas with low socio-economic status and/or a high share of immigrants (e.g., Hamburg and Saxony-Anhalt).

[23]I show later that the results are robust to the inclusion of state-by-year fixed effects, which can control for a possible differential roll-out of ADSPs across states.

personal resources. Concerning personal resources, anecdotal evidence suggests that the staff of a school often hampered a school's application as staff members were not willing to bear the high costs associated with setting up and running an ADSP.[24] Lobbying by parents with strong work preferences was also unlikely to occur, since parents were usually not aware of the IZBB program (unless it was discussed at the school conference) and the acceptance of ADSPs was very low, at least at the beginning of the IZBB investment period (Hagemann 2009, Augustin-Dittmann 2010). Nevertheless, I cannot completely rule out the possibility that mothers or a high anticipated demand influenced the application decision of school officials.

Thus, to strengthen the credibility of the instrument, I present the results of five empirical tests, all of which support the exogeneity assumption. First, I test based on observable maternal and state-level characteristics whether IZBB investments allocated to primary schools were as good as randomly assigned across counties. To do so, I compare mothers who lived in counties that had not allocated IZBB investments to primary schools yet with mothers who lived in counties that received IZBB investments for the very first time. Since maternal characteristics that are linked to maternal (full-time) employment might have been immediately affected by the provision of IZBB investments, I compare mothers based on their characteristics in the previous year. In addition, I test for systematic differences in childcare supply for preschool-aged children at the county level.[25] Table 3.1 depicts the results of this "balancedness check". It shows that there were no differences between these two types of mothers and counties.

Second, I test whether the allocation of IZBB investments (and thus IZBB program intensity) was correlated with pre-existing trends in maternal (full-time) employment (cf. Duflo 2001). Figure 3B.1 in Appendix 3B illustrates that prior to the IZBB program maternal employment trends were similar across counties with low and high IZBB program intensity (figure a). On the contrary, maternal full-time employment trends declined at a somewhat faster rate in counties with low program intensity (figure b), potentially leading to an overestimation of the treatment effect in the intensive margin model.

[24] These costs comprise, for example, the ex-ante time investments required to develop a school concept and the additional hours which had to be provided by the staff in the afternoon if an ADSP was implemented.

[25] As yearly data on the provision of childcare for zero to three year-olds and three to six year-olds are only available at the county level since 2007, I can only compare counties based on their IZBB investment status after 2007. Data were provided by the Federal Institute for Research on Building, Urban Affairs and Spatial Development (BBSR Bonn 2015).

Table 3.1. Balancedness Check: Maternal and County-Level Characteristics

	Difference of means	S.E. Difference
Maternal level		
Employed	-0.025	(0.032)
Full-time employed[a]	0.039	(0.041)
Education (in years)	0.126	(0.180)
Experience (in years)	-0.086	(0.491)
Work preferences: missing	0.013	(0.015)
Work preferences: low	-0.000	(0.033)
Work preferences: medium	-0.000	(0.039)
Work preferences: high	-0.012	(0.025)
Use of other types of care: missing	0.083*	(0.048)
Use of other types of care: yes	0.002	(0.036)
County level (past 2007)		
Child care coverage rate (children aged 0 to 3)	0.028	(0.041)
Child care coverage rate (children aged 3 to 6)	-0.010	(0.046)

Notes: Difference estimates are obtained from linear regressions of pre-period characteristics on an indicator for mothers who lived in counties (at maternal level)/counties (at county level) that received IZBB investments for primary schools for the very first time. Standard errors are clustered at the county level. The sample for the balancedness check at the maternal level is restricted to mothers who lived in counties that had not received any IZBB investments for primary schools yet or received these investments for the very first time. The sample for the balancedness check at the county level is restricted to counties that had not received any IZBB investments for primary schools yet or received these investments for the very first time.

[a] The difference in pre-period full-time employment means is conditional on being employed.

* $p<0.1$.

Sources: SOEPv30 (2003-2009), BBSR Bonn (2015) & SPI NRW (2010), own estimates.

Third, I carry out an inclusion test, i.e. besides the ADSP indicator and the covariates that are used in the final specification, I include the instrument in a single-equation probit for maternal (full-time) employment and test whether the coefficient on the instrument is significant. Table 3B.2 in Appendix 3B indicates that neither for the extensive nor for the intensive margin model the estimated coefficient on the instrument is statistically significant. In both models, coefficient estimates for the covariates and the value of the log likelihood function barely change upon the inclusion of the instrument.

Fourth, I perform a placebo test by adding IZBB investments of the subsequent year to the set of controls of the maternal (full-time) employment equation in the bivariate probit model. By construction, IZBB investments in year $t+1$ should not have an impact on maternal (full-time) employment in year t. Table 3B.3 in Appendix 3B shows for both the extensive and the intensive margin model that the coefficient on subsequent IZBB investments is small and insignificant.

Finally, I perform an additional placebo test based on a sample of childless women to investigate whether the results, which are presented in the next section, could be driven by some other unobservable differences between counties that received many IZBB investments for primary schools and counties that did not receive any IZBB investments or only very few of these investments for primary schools. As indicated by Table 3B.4 in Appendix 3B, the reduced form probit estimates are close to zero and insignificant, suggesting that the results are unlikely to be driven by other unobservable differences across counties. Overall, all of these tests provide some suggestive evidence that the exclusionary restriction is likely to hold in the German context.

3.5 Results

In this section, I start by presenting some descriptive evidence on the potential effect of ADSPs for primary school-aged children on maternal labor supply. Then, I show the bivariate probit results and investigate treatment effect heterogeneity. Finally, I show that the results are robust to numerous alternative specifications.

3.5.1 Descriptive Results

Figure 3.3 depicts the share of mothers with primary school-aged children in Germany who made use of an ADSP for at least one of their primary school-aged children between 1997 and 2013. As illustrated in this figure, the use of ADSPs started to increase markedly after the launch of the IZBB program in 2003. While prior to 2003 only 8.6% to 12.6% of all mothers with primary school-aged children made use of an ADSP, this share more than doubled until 2009 (26.9% in 2009), and it continued to increase even after the end of the IZBB program.[26] In the first five years of the IZBB investment period the share of mothers who made use of an ADSP increased at a smaller rate than in the final years of the IZBB period. The low response rate of mothers at the beginning of the IZBB period is likely explained by the low acceptance of ADSPs when the IZBB program was launched (Hagemann 2009, Augustin-Dittmann 2010) and by the fact that it can be very time-consuming to set up new ADSPs if additional staff has to be hired.

[26] After the end of the IZBB investment program the Federal Ministry of Education and Research has continued to promote the expansion of ADSPs in Germany. Based on the accompanying program "Ideas that go beyond! Learning all day" (in German: Ideen für mehr! Ganztägig lernen) 4.3 million euros are invested each year.

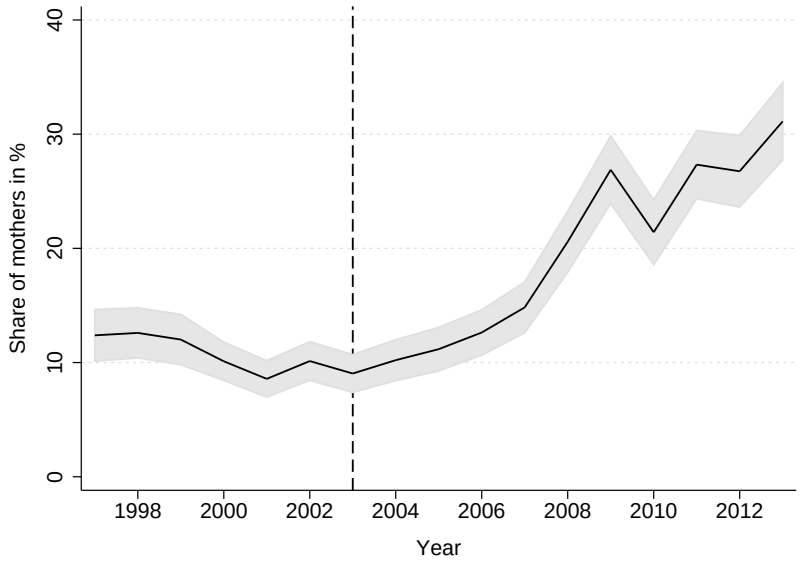

Figure 3.3. Maternal Use of All-Day Primary School Programs, Germany (1997 – 2013)

Notes: This figure shows the share of mothers with primary school-aged children in Germany who make use of an ADSP for at least one of their primary school-aged children. The gray shaded area represents the 90 percent confidence band. The sample encompasses 11,601 mother-year-observations. It excludes Bavarian and/or self-employed mothers, mothers who were in education and/or did not belong to the working age population, and mothers who had missing information in one of the dependent or explanatory variables. Estimates in a given year are based on at least 507 observations.

Source: SOEPv30 (1997-2013), own calculations.

Figure 3.4 illustrates the evolution of the employment share (figure a) and the full-time employment share (figure b) for mothers with primary school-aged children in Germany by maternal use of ADSPs.[27] It shows that, despite strong convergence over time, in almost all years maternal (full-time) employment shares were significantly larger among mothers who used an ADSP for primary school-aged children than among mothers who did not. However, the comparison of (full-time) employment shares between these two groups of mothers in a given year is little informative about the effect of ADSPs on maternal labor supply as mothers who select into ADSPs likely differ from mothers who do not.

In contrast, a relative comparison of the (full-time) employment shares of these two groups of mothers over time provides some initial evidence on the impact of ADSPs on maternal labor supply because of the exogenous variation induced by the IZBB investment program. The increasing employment share among mothers who did not make use

[27]In the pre-IZBB period and at the beginning of the IZBB investment period the (full-time) employment shares of mothers who made use of an ADSP for primary school-aged children are estimated based on less than 100 observations per year. Therefore, the estimates are imprecisely estimated and relatively volatile over time. The huge drop in the employment share of mothers who made use of an ADSP for primary school-aged children in 2004 likely represents an outlier.

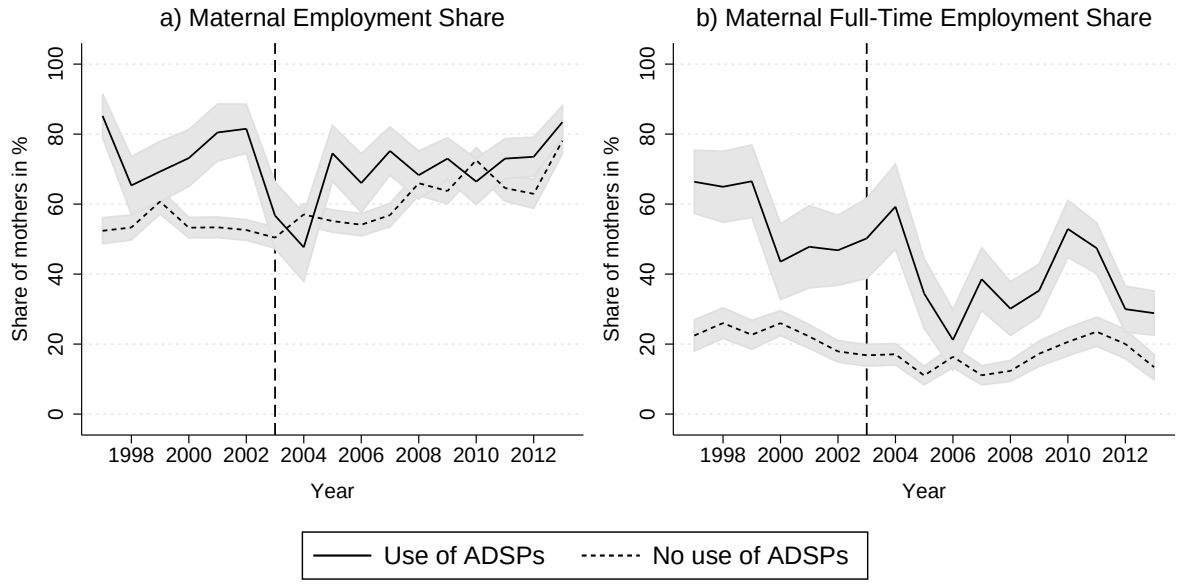

Figure 3.4. Maternal Employment and Full-Time Employment by Maternal Use of All-Day Primary School Programs, Germany (1997 – 2013)

Notes: Figure a) shows the share of employed mothers on all mothers with primary school-aged children in Germany who (i) make use of an ADSP for at least one of their primary school-aged children and (ii) do not make use of ADSPs for any of their primary school-aged children. For this figure the sample of mothers corresponds to the one of Figure 3.3. Figure b) shows the share of full-time employed mothers on all employed mothers with primary school-aged children in Germany again by maternal use of ADSPs for primary school-aged children. For this figure the sample of mothers corresponds to the one of Figure 3.3, except that it excludes non-employed mothers. Hence, estimates are based on 6,986 mother-year-observations. The gray shaded areas show 90 percent confidence bands.

Source: SOEPv30 (1997-2013), own calculations.

of an ADSP after the launch of the IZBB program suggests in combination with the non-declining employment share among mothers who made use of an ADSP, for instance, that ADSPs are very likely to have a positive impact on maternal employment. Because of the expansion of ADSPs, some mothers who were non-employed started to use an ADSP and were able to resume employment. As illustrated in Figure 3.4b, however, the majority of these mothers started to engage in part-time employment. The resulting composition effect is reflected by the declining share of full-time employed mothers among employed mothers who made use of an ADSP.[28] Due to the unknown size of the composition effect, it is impossible to assess whether ADSPs increased maternal full-time employment.

To ex ante rule out the possibility that the expansion of ADSPs crowded out alternative forms of care for primary school-aged children in Germany and, hence, had a limited

[28]Note that this decline is unlikely to be driven solely by general employment trends as the share of full-time employed mothers among employed mothers who did not make use of an ADSP stayed more or less constant over time. Given that there is no kink or jump in the trend of full-time employed mothers on employed mothers who used ADSPs in 2003, however, I cannot rule out the possibility of general employment trends based on this figure.

impact on maternal labor supply, I also analyze time trends of the shares of primary school-aged children who made use of ADSPs, horts[29], alternative forms of paid care (e.g., child minder), and unpaid care (e.g., relatives or friends) between 1997 and 2013. The results are depicted in Figure 3B.2 in Appendix 3B. As illustrated, there is no evidence for crowding-out of alternative types of care for primary school-aged children. After the launch of the IZBB program in 2003, the shares of primary school-aged children who made use of horts, paid care, and unpaid care did not decrease over time.[30]

Finally, to get a first impression of how much of the observed difference in (full-time) employment probabilities between mothers who make use of an ADSP for primary school-aged children and mother who do not make use of such a program is explained by selectivity of schools with ADSP and selection into ADSPs, I compare these two groups of mothers based on their observable characteristics. Summary statistics by maternal use of ADSP are presented in Table 3.2. Survey weights are used in order to obtain representative statistics for the specific sample of mothers. Standard errors are clustered at the county level to account for the panel structure of the data and the fact that mothers within the same county are more likely to share the same institutional context.

Table 3.2 shows that mothers who make use of ADSPs are 10.5 percentage points more likely to be employed, when compared to mothers who do not make use of ADSPs. Moreover, if employed these mothers are roughly two and a half times more likely to be full-time employed than their counterparts who do not make use of an ADSP. As expected, mothers who use ADSPs are more likely to live in a county that received IZBB investments for primary schools, less likely to live in a county that never received IZBB investments for primary schools between 2003 and 2009, and substantially more likely to live in a county with high cumulative IZBB investments per primary school. A comparison of maternal and family characteristics between these two groups of mothers demonstrates that mothers with primary school-aged children who make use of an ADSP are more likely

[29] A hort is a specific type of after-school care center in Germany. It is run by the institution of child and youth welfare and mainly attended by primary school students up to grade four. Despite horts are distinct from schools with ADSP, horts often closely cooperate with schools at the primary education level. Therefore, I separately show the share of primary school-aged children who made use of horts and other types of paid care, respectively.

[30] The sharp increase in the share of primary school-aged children who made use of horts in 2008/09 can be explained by a change in the questionnaire in 2009. Prior to 2009 information on the attendance of schools and center-based care was collected in the same question. As "after-school care center" was only one of many possible answers most of which referred to the attendance of different school types, it is very likely that parents were more likely to indicate the attendance of a hort after the introduction of two distinct questions, one for the attendance of schools and one for the attendance of horts.

Table 3.2. Summary Statistics by Maternal Use of ADSPs

	Use of ADSPs		No use of ADSPs	
	Mean	(S.E.)	Mean	(S.E.)
Main outcome variables				
Employed	0.677	(0.031)	0.572	(0.017)
Full-time employed	0.358	(0.041)	0.145	(0.015)
Instruments				
County received IZBB investments for primary schools	0.742	(0.045)	0.591	(0.033)
County never received IZBB investments for primary schools (2003 to 2009)	0.053	(0.018)	0.223	(0.034)
Cumulative IZBB investments per primary school/10 000 (in €)	10.2	(1.5)	3.7	(0.5)
Other explanatory variables				
Age (in years)	36.5	(0.4)	37.9	(0.2)
Migration background	0.225	(0.037)	0.315	(0.022)
Married	0.606	(0.045)	0.832	(0.015)
Education (in years)	12.5	(0.2)	11.9	(0.1)
Experience (in years)	10.1	(0.4)	10.1	(0.3)
Experience squared	138.2	(9.6)	142.4	(6.3)
Work preferences: missing	0.020	(0.010)	0.024	(0.006)
Work preferences: low	0.185	(0.031)	0.326	(0.020)
Work preferences: medium	0.614	(0.035)	0.523	(0.020)
Work preferences: high	0.181	(0.025)	0.127	(0.012)
Number of children	1.8	(0.1)	2.1	(0.0)
Young children present	0.251	(0.035)	0.288	(0.015)
Use of other care: missing	0.162	(0.023)	0.180	(0.007)
Use of other care: yes	0.476	(0.036)	0.245	(0.013)
Monthly non-wife net household income/100 (in €)	9.1	(0.4)	11.0	(0.2)
Area: rural	0.253	(0.061)	0.413	(0.031)
Area: urbanized	0.276	(0.055)	0.344	(0.030)
Area: urban	0.471	(0.084)	0.243	(0.035)
Female unemployment rate	0.090	(0.004)	0.066	(0.002)
Female part-time employment rate	0.132	(0.004)	0.121	(0.002)
Observations	693		4,323	

Notes: Standard errors are clustered at the county level. Full-time employment means are conditional on being employed.

Sources: SOEPv30 (2003-2009), BBSR Bonn (2015), ReGENESIS (2016) & SPI NRW (2010), own estimates.

to exhibit characteristics that are positively associated with maternal labor supply: On average, they are slightly younger, less likely to have a migration background and to be married, slightly better educated, and tend to have stronger work preferences. Moreover, they are less likely to have more children or a preschool-aged child, and more likely to live in urban areas. However, they tend to live in counties with slightly higher female unemployment and part-time employment rates. Except for the differences in experience and the probability of having young children, all differences in characteristics are statistically significant at the 5% level. Given this substantial positive selection into ADSPs based on

observable maternal characteristics, smaller effects for ADSPs on maternal labor supply are expected after accounting for selection on observables.

3.5.2 Multivariate Results

Table 3.3 presents the bivariate probit estimates that result from joint estimation of the maternal labor supply and ADSP use equations. Estimates for the extensive margin model are reported in columns 1 and 2, while estimates for the intensive margin model, which is conditional on maternal employment, are reported in columns 3 and 4. Standard errors are again clustered at the county level. The bivariate probit estimates suggest that voluntary ADSPs for primary school-aged children in Germany positively affect maternal employment, but have no effect on maternal full-time employment (first row columns 2 and 4). After accounting for selection on unobservables, mothers with primary school-aged children who make use of an ADSP have, on average, a 25.2 percentage points higher likelihood of being employed than their counterparts who do not make use of these programs. For the intensive margin, estimates suggest a negative, albeit insignificant effect of ADSPs on maternal full-time employment.[31]

As expected, the estimated coefficient on the instrument is strongly significant and positive in both models (second row columns 1 and 3), suggesting that an increase in cumulative IZBB investments per primary school increases the maternal probability of using an ADSP. If €10,000 more in a county are invested per primary school, this boosts the probability of using an ADSP for mothers with primary school-aged children in this county by 0.2 to 0.3 percentage points in the current and each of the three to four subsequent years, on average.[32] Although this effect seems to be small when compared to the effects of other covariates, it is considerable given that it accumulates and that before the

[31]The negative sign of the ADSP coefficient is somewhat counterintuitive. Yet, ADSPs for primary school-aged children may have a negative effect on maternal full-time employment probabilities if mothers substitute from care provided by family members (most prevalent form of care in Germany) to ADSPs and try to compensate this substitution by a reduction in working hours in order to regain social esteem. Such a substitution will not be visible in the crowding-out analysis (see Figure 3B.2 in Appendix 3B) if family members continue to provide care, since the SOEP data lack information on the hours of care.

[32]Using cumulative IZBB investments, the underlying dynamics are not modeled. To investigate the underlying dynamics, I estimated bivariate probit models in which I used current and lagged IZBB investments per primary school (up to five lags) rather than cumulative IZBB investments per primary school as instruments (results available upon request). I find that the effect of a €10,000 IZBB investment on maternal use of ADSPs gradually decays over time (0.5 percentage points in the year of investment to 0.1 percentage points in the third year after investment). High lags are imprecisely estimated, however. The cumulative effect represents the average effect across the current and subsequent years.

Table 3.3. Bivariate Probit Estimates of Maternal Labor Supply Models

	Extensive margin		Intensive margin[a]	
	Use of ADSPs	Employed	Use of ADSPs	Full-time employed
Use of ADSPs		1.111***		-0.291
		(0.243)		(0.448)
Cumulative IZBB investments per primary school/10 000	0.015***		0.020***	
	(0.006)		(0.006)	
Age (in years)	-0.039***	-0.108***	-0.046**	-0.078***
	(0.013)	(0.010)	(0.019)	(0.015)
Migration background	0.175	0.023	0.169	0.503***
	(0.139)	(0.075)	(0.135)	(0.123)
Married	-0.341***	0.362***	-0.344***	-0.282**
	(0.090)	(0.087)	(0.110)	(0.119)
Education (in years)	0.069***	0.144***	0.077***	0.146***
	(0.016)	(0.016)	(0.019)	(0.023)
Experience (in years)	0.032*	0.236***	0.008	0.075***
	(0.019)	(0.016)	(0.026)	(0.027)
Experience squared	-0.000	-0.003***	0.000	0.000
	(0.001)	(0.001)	(0.001)	(0.001)
Number of children	-0.177***	0.030	-0.331***	-0.150*
	(0.054)	(0.043)	(0.071)	(0.081)
Young children present	0.092	-0.673***	0.322***	-0.058
	(0.082)	(0.062)	(0.105)	(0.111)
Monthly non-wife net household income/100	-0.024***	-0.035***	-0.017**	-0.058***
	(0.007)	(0.006)	(0.007)	(0.010)
Area: urbanized	0.340***	-0.082	0.277**	0.039
	(0.097)	(0.081)	(0.122)	(0.118)
Area: urban	0.672***	-0.191*	0.635***	0.039
	(0.126)	(0.103)	(0.172)	(0.166)
Female unemployment rate	-2.791	1.187		
	(2.567)	(2.088)		
Female part-time employment rate			1.244	5.339***
			(1.665)	(1.644)
ρ	-0.521***		0.380	
	(0.137)		(0.249)	
State FE	Yes		Yes	
Year FE	Yes		Yes	
ATE	0.252***		-0.061	
	(0.048)		(0.087)	
Value of log likelihood	-3,655.69		-2,131.91	
Observations	5,016		3,050	

Notes: Coefficient estimates. Models additionally include a constant. Standard errors are clustered at the county level and are reported in parentheses.

[a] Bivariate probit estimates for the intensive margin are conditional on maternal employment.

* p<0.1 ** p<0.05 *** p<0.01.

Sources: SOEPv30 (2003-2009), BBSR Bonn (2015), ReGENESIS (2016) & SPI NRW (2010), own estimates.

launch of the IZBB program only 10% of all mothers with primary school-aged children made use of an ADSP.

Turning to the estimates of the correlation coefficient ρ, the bivariate probit estimate for the extensive margin model indicates a strong negative correlation between the structural errors of the employment and ADSP use equations. As discussed in section 3.4.1, negative selection into ADSPs may result from the fact that mothers with stronger preferences for ADSPs exhibit unobservable characteristics which make them less likely to engage in employment. Beyond that, it is very likely that mothers with more favorable unobservable labor market characteristics have better (financial and social) resources to promote their child's interests (cf. Börner et al. 2010). Using alternative types of (high-quality) childcare, these mothers would be able to work even in the absence of ADSPs.

In contrast to the extensive margin model, the estimate of ρ is positive and insignificant for the intensive margin model. The different sign of the correlation coefficient in this model is likely driven by the fact that the sample of employed mothers is a selective sample. Within this sample, full-time employed mothers are very likely to exhibit stronger work preferences than part-time employed mothers. Moreover, full-time employed mothers probably have a higher likelihood of using paid and public care arrangements such as ADSPs, since it is more difficult to arrange family care for a full than a half day. Hence, even if full-time employed mothers were to more strongly oppose ADSPs (cf. Börner et al. 2010), given their stronger work preferences, they would use ADSPs for primary school-aged children more often than part-time employed mothers.

Apart from selection on unobservables, there is also some evidence for selection on observables. In line with the descriptive results, I find that (employed) mothers with primary school-aged children in Germany exhibit a higher probability of using an ADSP if they are younger, non-married, better educated, and have more labor market experience (only significant in the extensive margin model). Moreover, they have a higher propensity of using ADSPs for primary school-aged children if they have fewer children or preschool-aged children (only significant in the intensive margin model), if their family exhibits a lower monthly non-wife net household income, and if they live in non-rural areas. Many of these characteristics are also positively linked to maternal labor supply at both margins, suggesting positive selection on observables.

Estimated coefficients on covariates exhibit the expected sign for all but two covariates, whose coefficient signs seem to be counterintuitive at first sight. First, being married is positively linked to maternal employment. Because of the so-called "Ehegattensplitting" (a special tax system for spouses) and the co-insurance of dependents, in Germany married mothers have much lower work incentives than non-married mothers. However, unlike non-married mothers, married mothers can share care responsibilities with their spouses, which eases maternal employment. Therefore, and because of the cut in unemployment benefits in 2005, married mothers often decide to work marginally. Marginal employment does not affect the tax advantages of spouses, the co-insurance of dependents, and the amount of unemployment benefits. As I classified mothers who indicated marginal or irregular employment (roughly 11% of all mother-year-observations) as employed mothers, this explains part of the large and significantly positive impact of being married on maternal employment.[33] Second, the estimates for the extensive margin model indicate that living in urban relative to rural areas decreases the probability of maternal employment. Due to the greater availability of jobs in urban areas, the sign of this coefficient seems to be implausible at first sight, but taking into consideration that in rural areas grandparents are more likely to live in the same household or in short distance, the negative effect of community size on maternal employment is likely explained by better access to family care.[34]

3.5.3 Treatment Effect Heterogeneity

Table 3.4 shows the bivariate probit estimates for the extensive margin model by maternal education. A comparison between mothers with a vocational degree or lower and mothers with a university degree illustrates that the large effect of ADSPs on maternal employment is concentrated among mothers with at most a vocational degree. On average, the use of an ADSP for primary school-aged children increases the employment probability

[33]Note that the magnitude of the estimated coefficient substantially decreases in a model that classifies marginally employed mothers as non-employed mothers. Yet, the estimated marital status coefficient continues to be significantly positive at the 5% level. This suggests that shared care duties between spouses increase the probability of maternal employment.

[34]For my sample of mothers, between 2003 and 2009, the probability of living in a multi-generational household (i.e., a household where at least one grandparent lives with high probability) was almost twice as large in rural when compared to urban areas. Moreover, the share of mothers who made use of family care for their primary school-aged children was roughly five percentage points higher in rural relative to urban areas (rural areas: 25.4%).

Table 3.4. Treatment Effect Heterogeneity by Maternal Education

Estimates of	Vocational degree or lower	University degree
ADSP coefficient in	1.311***	-0.046
employment equation	(0.234)	(0.735)
Coefficient on instrument	0.012**	0.028**
in ADSP equation	(0.006)	(0.011)
ρ	-0.628***	0.153
	(0.128)	(0.392)
ATE	0.297***	-0.011
	(0.045)	(0.170)
Observations	4,173	843

Notes: Bivariate probit estimates for the extensive margin model. Cumulative IZBB investments per primary school at the county level are used as instrument. Covariates correspond to those of Table 3.3, except that the years of schooling are no longer included in the bivariate probit regressions. Standard errors are clustered at the county level and are reported in parentheses.
** p<0.05 *** p<0.01.
Sources: SOEPv30 (2003-2009), BBSR Bonn (2015), ReGENESIS (2016) & SPI NRW (2010), own estimates.

of these mothers by 29.7 percentage points. The remaining coefficient estimates are consistent with the findings presented in the previous section. In particular, a larger amount of cumulative IZBB investments per primary school raises the probability of ADSPs use for both groups of mothers. Moreover, the estimated correlation coefficients are significantly negative for mothers with at most a vocational degree and slightly positive but insignificant for mothers with a university degree, which matches the correlation structure uncovered for the extensive and intensive margin models when pooling all mothers. This finding is intuitive as mothers with a university degree often correspond to employed mothers, but only make up a small share of the full sample of mothers. In line with expectations, I do not find any evidence for a positive effect of ADSPs on maternal full-time employment for both groups of mothers.

One likely explanation for this observed differential impact of ADSPs on maternal labor supply at the extensive margin by education are different working contracts offered to high- and low-skilled workers in Germany. In contrast to low-skilled jobs, high-skilled jobs are often characterized by a high degree of task specialization. Therefore, acquiring task-specific knowledge is an important prerequisite for the successful completion of tasks in high-skilled positions. As the acquisition of this knowledge can be very time-consuming, it is less costly (and more efficient) for firms to offer full-time jobs to high-skilled workers. On the contrary, part-time jobs and jobs that allow for more flexible working hours are

more often offered to low-skilled workers. Given that in Germany ADSPs for primary school-aged children, although increasing the time spent in schools, generally remain incompatible with regular full-time working schedules, mothers with a university degree are rather unlikely to respond to ADSPs by adjusting their labor supply at the extensive margin. Having stronger work preferences and better financial resources than lower educated mothers, mothers with a university degree possibly send their child to paid childcare and engage in employment (most likely full-time employment) even in the absence of ADSPs. However, with the expansion of ADSPs these mothers likely substitute some hours of expensive paid care by relatively cheaper ADSP care and continue using paid care arrangements after the ADSP finishes in the afternoon. This crowding-out of paid care arrangements remains invisible in the SOEP data, since they lack information on the hours of care provided by private care providers.

Apart from heterogeneity by education, I explored treatment effect heterogeneity along other dimensions: marital status, presence of preschool-aged children in the household, age of youngest primary school-aged child in the household, and region of residence (East vs. West Germany). I did not find any robust evidence for treatment effect heterogeneity along these dimensions.

3.5.4 Robustness Checks

This section provides numerous robustness checks. More specifically, I test the sensitivity of results to alternative estimation methods and coding decisions made throughout the data preparation process of the IZBB investment data set, to the use of lagged cumulative IZBB investments or current and lagged year-specific IZBB investments, to variations in the definitions of the maternal ADSP use and employment indicators, and to the inclusion of additional or alternative sets of control variables. For the sake of brevity, I mainly focus on the sensitivity of results for the extensive margin model. I find that the results are robust to all but one variation. In the extensive margin model, the 2SLS estimate of the ADSP coefficient turns out to be insignificant. This is, however, not very surprising given that 2SLS estimates are less efficiently estimated in the presence of selection on unobservables and uncover a different treatment effect under treatment effect heterogeneity. The results of the robustness checks are shown in Appendix 3B and discussed in more detail below.

Table 3B.5 presents the estimation results which are obtained from 2SLS. Coefficient estimates almost always exhibit the same sign as in the bivariate probit model. The coefficient estimates which exhibit a different sign (number of children and female county-level unemployment rate in column 2) are either close to zero or imprecisely estimated and, hence, insignificant. The estimated coefficients on the instrument (second row in columns 1 and 3) are again positive and highly significant. For both the extensive and the intensive margin model, the relevance of the instrument is further supported by a first-stage F-statistic which is larger than ten. As in the bivariate probit model, the effect of ADSPs on maternal labor supply at the intensive margin (first row in column 4) is negative and insignificant. The effect of ADSPs on maternal labor supply at the extensive margin is again positive, but much smaller in magnitude and more imprecisely estimated (first row in column 2). Therefore, under 2SLS I no longer find a significantly positive effect of ADSPs for primary school-aged children on maternal employment.

Given that in applied research 2SLS and bivariate probit estimates often coincide, one common belief among many applied researchers is that 2SLS and bivariate probit estimations should produce very similar results. Yet, Chiburis et al. (2012) show in a simulation study that bivariate probit and 2SLS estimates can substantially differ in samples of up to 5,000 observations and/or if treatment probabilities are close to zero or one, both of which is the case in this study.[35] From a theoretical perspective, this belief is also causeless because bivariate probit estimates and 2SLS estimates identify different treatment effects under treatment effect heterogeneity. In section 3.5.3 some evidence for treatment effect heterogeneity was provided, thus, making it very plausible that bivariate probit and 2SLS results differ in this study. However, even if the magnitude of the estimated treatment effect is similar for both estimation methods, it occasionally occurs that bivariate probit estimates are significant while 2SLS are not. This is because bivariate probit estimates are more efficiently estimated if the underlying distributional assumption is correct. In the case of this study, standard errors of the treatment effect are smaller under bivariate probit than 2SLS estimation. This provides an additional explanation for the insignificance of the ADSP coefficient that is obtained by 2SLS.

As an additional robustness check for the results of the extensive margin model, I computed a lower bound estimate of the ATE using a single-equation probit for maternal

[35]For the extensive margin model, the sample size actually slightly exceeds 5,000 observations. Yet, depending on the value of ρ, large deviations between bivariate probit and 2SLS estimates can even exist in samples of more than 5,000 observations (cf. Chiburis et al. 2012).

employment. After controlling for differences in observable maternal characteristics, a single-equation probit for maternal employment provides a lower bound estimate for the ATE because of the negative correlation between the two structural errors in the bivariate probit model for the extensive margin. The probit estimates show that this lower bound estimate of the ATE is roughly five times smaller than the bivariate probit estimate of the ATE. However, the lower bound ATE estimate remains statistically significant at the 1% level. Thus, this finding further supports the conclusion that ADSPs for primary school-aged children have a positive impact on maternal labor supply at the extensive margin.[36]

Finally, I turn to robustness checks that test the sensitivity of results to alternative specifications of the bivariate probit model for the extensive margin. Table 3B.6 shows that the results are robust to variations in the specification of the instrument. In column 2, I remove mothers who lived in counties with primary schools that received implausible low investment amounts in a given year to show that results are robust to reporting errors in the IZBB investment data set. In column 3, I remove mothers from Thuringia to demonstrate that the results are robust to the imputation strategy used to impute cumulative IZBB investment amounts for counties in Thuringia. In column 4, I use cumulative IZBB investments that are lagged by one year to test whether lagged instead of current cumulative IZBB investments per primary school impact the maternal use of ADSPs. As lagged cumulative IZBB investments ignore the immediate effect of IZBB investments, however, I decided to use the current amount of cumulative IZBB investments per primary school as instrument in the main specification. In column 5, I use current and lagged year-specific IZBB investments per primary school instead of cumulative IZBB investments as instruments. This specification demonstrates that lagged investments predict maternal use of ADSPs, even conditional on current investments, thereby supporting the use of cumulative rather than current IZBB investments.

Table 3B.7 indicates that the results are robust to alternative definitions of the maternal ADSP use indicator. As information on the attendance of ADSPs was collected at the child level and some mothers in the sample had several primary school-aged children for some of whom they used ADSPs in a given year and for some of whom they did not, I had to aggregate the information from the child level at the maternal level.[37] When compared

[36] This finding is robust to the estimation of a linear probability model.

[37] 15.3% of the mother-year-observations refer to mothers with more than one primary school-aged child.

to the results under the main definition of the ADSP use indicator (column 1), the results under the second more conservative definition of the ADSP use indicator (column 2) are almost identical. This suggests that mothers who had more than one primary school-aged child, in general, made the same decisions for their primary school-aged children.[38] In column 3, I disregard mother-year-observations of mothers with several children that attend primary school in a given year. Thus, for this sample of mothers the information of ADSP use directly maps from children to mothers. I continue to find a significantly positive effect of voluntary ADSPs on maternal employment, but the coefficient estimate is roughly five percentage points smaller than in the main specification.

Table 3B.8 demonstrates that the results are robust to alternative definitions of the maternal employment indicator. Given that with the Hartz IV reform in 2005 marginal employment became much more common in Germany, I investigate whether results continue to hold if I reclassify marginally employed mothers as non-employed mothers (column 2) or if I remove these mothers from the sample (column 3). When compared to the ATE of the main specification (column 1), the ATE after reclassification of marginally employed mothers is larger. This finding is plausible because the German federal government introduced marginal employment in order to increase job finding rates among the unemployed. The ATE is of similar magnitude after removing marginally employed mothers from the sample.

Table 3B.9 indicates the bivariate probit estimates for the extensive margin model after including additional sets of control variables. It shows that the results are robust if I control for the use of alternative types of care for primary school-aged children (columns 1 and 2), maternal work preferences (column 3), or state-by-year fixed effects rather than state and year fixed effects (column 4). Access to alternative types of care for primary school-aged children likely affects both maternal use of ADSPs for primary school-aged children and maternal labor supply. However, likewise ADSPs use, the use of alternative types of care for primary school-aged children is endogenous, hence, calling for an additional instrument and the estimation of a more complex trivariate probit. Therefore, and because no data on paid and unpaid care for primary school-aged children were collected in 2003, I decided against controlling for alternative types of care in the main specification. However, if I were to control for the use of alternative types of care in a bivariate

[38] In fact, only 3% of the mother-year-observations of mothers with several primary school-aged children refer to mothers who used ADSPs for some of their primary school-aged children but not for others.

model, both the indicator for the use of alternative types of care and the indicator for missing care information are significant in both equations. The ATE changes very little, independent of whether I keep observations of the year 2003 (included in the missing dummy, column 1) or whether I exclude them (column 2).

Unobserved differences in work preferences are one of the major concerns generating positive selection into ADSPs. Therefore, I proxy work preferences by information on the importance of job success, which was gathered on an irregular basis in 1995, 2004, 2008, and 2012 in the SOEP. In order to impute the information in the remaining survey years, I assume that work preferences are stable over time and use the same information for years following each survey year.[39] Moreover, I include a dummy for missing work preferences due to item-nonresponse in the bivariate probit regressions. Although I find that mothers with low work preferences are significantly less likely to use ADSPs for primary school-aged children and to engage in employment than mothers with high work preferences, the bivariate probit estimate of the ATE only decreases marginally upon the inclusion of work preferences in the model. In line with the strong negative estimated correlation coefficient, this finding suggests that selection based on maternal work preferences is of minor importance in the German context. Therefore, and because of the imputation requirement, which relies on the assumption of stable work preferences over time, I decided against controlling for work preferences in the main specification.

Table 3B.10 illustrates that the results are robust to the modification or removal of possible endogenous controls. In column 1, I employ the education of a partner or spouse rather than need-weighted deflated non-wife net household income to control for (financial) resources of the household, since labor supply within a household might be codetermined. As non-wife net household income better captures the availability of financial resources and results are robust to this variation, I decided to control for non-wife net household income in the main specification, however. In column 2, I remove the female regional unemployment rate because county-level female labor market outcomes embed those of mothers with primary school-aged children. As official labor market indicators for distinct groups of females (e.g., childless women) are not reported at the county level, however, I had to use the overall county-level female unemployment rate in the main specification to control for county-level labor market trends.

[39] I also use this information for antecedent years if survey respondents entered the SOEP due to sample refreshment.

3.6 Discussion

In this study, I analyzed the effect of voluntary ADSPs for primary school-aged children on maternal labor supply. I focused on Germany, a country that has strongly expanded the supply of ADSPs since 2003. In order to account for selectivity of primary schools with ADSP and selection into ADSPs, I estimated bivariate probit models. I exploited regional and temporal variation in IZBB investments allocated to primary schools in order to identify these models. The key finding is that voluntary ADSPs for primary school-aged children increase maternal labor supply at the extensive margin, while they have no effect on maternal labor supply at the intensive margin.[40] This finding is in line with expectations as ADSPs in Germany, albeit increasing the time spent in primary schools, generally remain incompatible with regular full-time working schedules.

The incompatibility between ADSPs and full-time working schedules in Germany is additionally reflected in different labor supply responses of mothers with primary school-aged children by education. My results show that the large effect of voluntary ADSPs at the extensive margin is concentrated among mothers with at most a vocational degree.[41] Facing more flexible work arrangements than mothers with a university degree, these mothers can easily combine work and family life after getting access to this highly subsidized type of childcare. On the contrary, mothers with a university degree continue to face substantial difficulties in combining work and family life even after the expansion of ADSPs in Germany, since high-skilled positions are often offered as full-time positions. Thus, to bridge the gap between working hours and ADSP schedules for these mothers, it would be necessary to extend the operating hours of ADSPs until 6 p.m.

Finally, my results reveal that, in the German context, selection into ADSPs for primary school-aged children is negative. This suggests that mothers with less favorable unobservable labor market characteristics are more likely to use ADSPs than mothers with more favorable unobservable labor market characteristics. The relatively low costs

[40] Note that a causal interpretation of the results for the intensive margin is only possible under very strong assumptions. This also holds true if I were to estimate a trivariate probit instead of two bivariate probit models (cf. Angrist 2001, Staub 2014). Yet, given that it is very likely that ADSPs in Germany only have a causal impact on maternal labor supply at the extensive margin, this limitation is of minor importance for this study.

[41] Compared to mothers with a university degree, mothers with at most a vocational degree were roughly 20 percentage points less likely to be employed pre-investment (employment share of roughly 50%). This low level of initial employment may explain why the effects uncovered in this study exceed those usually uncovered in the literature (e.g., Cascio et al. (2015) for a survey).

of ADSPs in combination with the broad supply of social and cultural activities, which render voluntary ADSPs particularly attractive to families with low socio-economic status and tighter budget constraints, likely provide an explanation for this finding.

Overall, the results of this study show that a current ADSP's capability to activate the unused labor force potential of mothers with primary school-aged children is still limited to specific groups of mothers, possibly due to the restrictive operating hours of ADSPs. Therefore, it is very likely that in Germany the expansion of ADSPs had a smaller effect on maternal labor supply than desired by politicians. Moreover, given the different responses by maternal education, it is evident that in Germany the expansion of ADSPs likely had some unintended consequences. As ADSPs for primary school-aged children mainly increased the part-time employment probability of low-skilled mothers, there is a high chance that the expansion of ADSPs increased gender inequality in working hours and wages. Thus, if politicians aim at further promoting maternal labor supply and improving gender equality, one important next step of the policy agenda could be the extension of ADSP schedules to better match regular full-time working schedules in Germany. However, not only the extension of ADSP schedules is important. Given that ADSPs are generally not offered during public school holidays, another challenging task for politicians is the provision of care during school holidays.

The sole focus of this study was on maternal labor supply responses to ADSPs. As ADSPs may also have far-reaching consequences for other family members, two additional questions need to be addressed by future research: First, how do ADSPs affect the labor supply of spouses, i.e. do ADSPs lead to a more equal labor supply allocation within the household? Second, do ADSPs improve child outcomes? Given that the achievement of more educational and social justice was one explicit goal of the expansion of ADSPs in Germany, it would particularly be interesting to look at children of socio-economically disadvantaged families. Yet, to rule out a detrimental impact of ADSPs on children, in general, it would be indispensable to look at children of other families, as well. Taking into account this study's findings and the suggested policy implications, in particular, the negation of a detrimental impact of ADSPs on children would be of great importance.

References

Angrist, J. D. (2001): "Estimation of Limited Dependent Variable Models with Dummy Endogenous Regressors: Simple Strategies for Empirical Practice," *Journal of Business & Economic Statistics*, 19, 2–16.

Augustin-Dittmann, S. (2010): "The Development of All-Day Schooling in Germany: How Was it Possible in this Conservative Welfare State?" *German Policy Studies*, 6, 49–81.

Baker, M., J. Gruber, and K. Milligan (2008): "Universal Child Care, Maternal Labor Supply, and Family Well-Being," *Journal of Political Economy*, 116, 709–745.

Bauernschuster, S. and M. Schlotter (2015): "Public Child Care and Mothers' Labor Supply - Evidence from Two Quasi-Experiments," *Journal of Public Economics*, 123, 1–16.

Beblo, M., C. Lauer, and K. Wrohlich (2005): "Ganztagsschulen und Erwerbsbeteiligung von Müttern - eine Mikrosimulationsstudie für Deutschland (All-Day Schools and Labour Force Participation of Mothers – A Micro-Simulation Study for Germany)," *Zeitschrift für Arbeitsmarktforschung - Journal for Labour Market Research*, 38, 357–372.

Becker, G. S. (1991): *A Treatise on the Family*, Cambridge, MA: Harvard University Press.

Berlinski, S. and S. Galiani (2007): "The Effect of a Large Expansion of Pre-Primary School Facilities on Preschool Attendance and Maternal Employment," *Labour Economics*, 14, 665–680.

Bertelsmann Stiftung (2012): *Ganztagsschule als Hoffnungsträger für die Zukunft? Ein Reformprojekt auf dem Prüfstand. Expertise des Deutschen Jugendinstituts (DJI) im Auftrag der Bertelsmann Stiftung*, Gütersloh: Verlag Bertelsmann Stiftung.

Berthelon, M. E., D. I. Kruger, and M. Oyarzùn (2015): "The Effects of Longer School Days on Mothers' Labor Force Participation," IZA Discussion Papers 9212, Institute for the Study of Labor (IZA).

Blau, D. M. and P. K. Robins (1988): "Child-Care Costs and Family Labor Supply," *The Review of Economics and Statistics*, 70, 374–81.

Blossfeld, H.-P., W. Bos, H.-D. Daniel, B. Hannover, D. Lenzen, M. Prenzel, H.-G. Rossbach, R. Tippelt, and L. Wössmann (2013): *Zwischenbilanz*

Ganztagsgrundschulen: Betreuung oder Rhythmisierung? Gutachten, Münster: Waxmann.

Börner, N., K. Beher, W. Düx, and I. Züchner (2010): "Lernen und Fördern aus Sicht der Eltern," in *Lernen und Fördern in der offenen Ganztagsschule. Vertiefungsstudie zum Primarbereich in Nordrhein-Westfalen*, Weinheim u.a.: Juventa, Juventa: Materialien, 143–225.

Bundesinstitut für Bau-, Stadt- und Raumforschung (BBSR Bonn) (2015): "INKAR - Indikatoren und Karten zur Raum- und Stadtentwicklung," www.bbsr.bund.de. Accessed on 22 May 2015.

Bundesministerium für Bildung und Forschung (Ed.) (BMBF) (2003): "Verwaltungsvereinbarung zwischen der Bundesrepublik Deutschland und den Bundesländern zum Investitionsprogramm 'Zukunft Bildung und Betreuung' 2003 - 2007," www.ganztagsschulen.org/_media/20030512_verwaltungsvereinbarung_zukunft_bildung_und_betreuung.pdf. Accessed on 12 September 2015.

——— (2009): "Gut angelegt – Das Investitionsprogramm Zukunft Bildung und Betreuung," www.ganztagsschulen.org/_media/gut_angelegt.pdf. Accessed on 12 September 2015.

Bundesministerium für Familie, Senioren, Frauen und Jugend (BMFSFJ) (1996): *Zeit im Blickfeld – Ergebnisse einer repräsentativen Zeitbudgeterhebung (Schriftenreihe Bd. 121)*, Stuttgart, Berlin, Köln: Kohlhammer.

——— (2011): "Vereinbarkeit von Familie und Beruf mit Schulkindern," www.bmfsfj.de/bmfsfj/service/publikationen/vereinbarkeit-von-familie-und-beruf-mit-schulkindern/86392. Accessed on 11 September 2015.

——— (2013): *13. Kinder- und Jugendbericht. Bericht über die Lebenssituation junger Menschen und die Leistungen der Kinder- und Jugendhilfe in Deutschland*, Berlin: Bundestagsdrucksache 17/12200, www.bmfsfj.de/RedaktionBMFSFJ/Broschuerenstelle/Pdf-Anlagen/13-kinder-jugendbericht,property=pdf,bereich=bmfsfj,rwb=true.pdf. Accessed on 11 September 2015.

Bundesvereinigung Kulturelle Kinder- und Jugendbildung e.V. (BKJ) (2006): "Kulturelle Bildung und Ganztagsschulen: Rahmenbedingungen und Umsetzung von Kooperationen in den Ländern. Eine Ländersynopse," www.miz.org/dokumente/bkj_Laendersynopse.pdf. Accessed on 12 September 2015.

Cascio, E. U. (2009): "Maternal Labor Supply and the Introduction of Kindergartens into American Public Schools," *Journal of Human Resources*, 44, 140–170.

Cascio, E. U., S. J. Haider, and H. Skyt Nielsen (2015): "The Effectiveness of Policies that Promote Labor Force Participation of Women with Children: A Collection of National Studies," *Labour Economics*, 36, 64–71.

Chiburis, R. C., J. Das, and M. Lokshin (2012): "A Practical Comparison of the Bivariate Probit and Linear IV Estimators," *Economics Letters*, 117, 762–766.

Contreras, D. and P. Sepúlveda (2017): "Effect of Lengthening the School Day on Mother's Labor Supply," *World Bank Economic Review*, 31, 747–766.

Dehos, F. and M. Paul (2017): "The Effects of After-School Programs on Maternal Employment," SOEPpapers on Multidisciplinary Panel Data Research 905, Deutsches Institut für Wirtschaftsforschung (DIW Berlin).

Deutsches Jugendinstitut (Ed.) (DJI) (2005): "Zahlenspiegel 2005. Kindertagesbetreuung im Spiegel der Statistik," `www.bmfsfj.de/doku/Publikationen/zahlenspiegel2005/01-Redaktion/PDF-Anlagen/Gesamtdokument,property=pdf,bereich=zahlenspiegel2005,sprache=de,rwb=true.pdf`. Accessed on 9 September 2015.

Dohmen, D. and K. Himpele (2006): "Kosten von Ganztagsschulen," *Recht der Jugend und des Bildungswesens*, 1/2006, 1–22.

Duflo, E. (2001): "Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment," *American Economic Review*, 91, 795–813.

Felfe, C., M. Lechner, and P. Thiemann (2016): "After-School Care and Parents' Labor Supply," *Labour Economics*, 42, 64–75.

Fitzpatrick, M. (2010): "Preschoolers Enrolled and Mothers at Work? The Effects of Universal Prekindergarten," *Journal of Labor Economics*, 28, 51–85.

Gambaro, L., J. Marcus, and F. Peter (2016): "School Entry, Afternoon Care and Mothers' Labour Supply," DIW Discussion Papers 1622, Deutsches Institut für Wirtschaftsforschung (DIW Berlin).

Gelbach, J. B. (2002): "Public Schooling for Young Children and Maternal Labor Supply," *American Economic Review*, 92, 307–322.

Hagemann, K. (2009): "Die Ganztagsschule als Politikum. Die bundesdeutsche Entwicklung in gesellschafts- und geschlechtergeschichtlicher Perspektive," in *Ganztägige Bildung und Betreuung*, ed. by L. Stecher, C. Allemann-Ghionda, W. Helsper, and E. Klieme, Weinheim: Beltz, 209–229.

Havnes, T. and M. Mogstad (2011): "Money for Nothing? Universal Child Care and Maternal Employment," *Journal of Public Economics*, 95, 1455–1465.

Konferenz der Kultusminister der Bundesrepublik Deutschland (Ed.) (KMK) (2014): "Allgemein bildende Schulen in Ganztagsform in den Ländern in der Bundesrepublik Deutschland, verschiedene Jahre (2008-2014)," `www.kmk.org/statistik/schule/statistische-veroeffentlichungen/allgemeinbildende-schulen-in-ganztagsform.html`. Accessed on 23 May 2015.

Kreyenfeld, M. and E. Geisler (2006): "Müttererwerbstätigkeit in Ost- und Westdeutschland," *Zeitschrift für Familienforschung - Journal of Family Research*, 18, 333–360.

Marcus, J., J. Nemitz, and K. C. Spiess (2013): "Ausbau der Ganztagsschule: Kinder aus einkommensschwachen Haushalten im Westen nutzen Angebote verstärkt," *DIW-Wochenbericht 2013*, 27, 11–23.

Marcus, J., J. Nemitz, and K. C. Spiess (2016): "Veränderungen in der gruppenspezifischen Nutzung von ganztägigen Schulangeboten – Längsschnittanalysen für den Primarbereich (Changes in the Group Specific Use of All-Day Schooling: Longitudinal Analyses for the Primary Education Sector)," *Zeitschrift für Erziehungswissenschaft - German Journal of Educational Science*, 19, 415–442.

Nemitz, J. (2015): "The Effect of All-Day Primary School Programs on Maternal Labor Supply," ECON Working Papers 213, Department of Economics, University of Zurich.

Rainer, H., S. Bauernschuster, W. Auer, N. Danzer, M. Hancioglu, B. Hartmann, T. Hener, C. Holzner, N. Ott, J. Reinkowski, and M. Werding (2013): "Kinderbetreuung," ifo Forschungsberichte 59, ifo Institut.

ReGENESIS (2016): "Statistik der Allgemeinbildenden Schulen (Tabelle: 21111kj001)," `http://regenesis.pudo.org/`. Accessed on 27 August 2016.

Ribar, D. C. (1992): "Child Care and the Labor Supply of Married Women: Reduced Form Evidence," *Journal of Human Resources*, 27, 134–165.

Shure, N. (2016): "School Hours and Maternal Labor Supply: A Natural Experiment from Germany," DoQSS Workings Papers 16-13, Department of Quantitative Social Science, UCL Institute of Education, University College London.

Sozialpädagogisches Institut NRW, TH Köln (SPI NRW) (2010): "Investitionsprogramm 'Zukunft Bildung und Betreuung' 2003-2009. Geförderte Schulen in den Jahren 2003, 2004, 2005, 2006, 2007, 2008, 2009, verschiedene Bundesländer," Wiesbaden.

Statistisches Bundesamt (2012): "Vereinbarkeit von Familie und Beruf. Ergebnisse des Mikrozensus 2011," Wiesbaden. www.destatis.de/DE/Publikationen/WirtschaftStatistik/Bevoelkerung/VereinbarkeitFamilieBeruf_122012.pdf?__blob=publicationFile. Accessed on 10 September 2015.

Staub, K. E. (2014): "A Causal Interpretation of Extensive and Intensive Margin Effects in Generalized Tobit Models," *Review of Economics and Statistics*, 96, 371–375.

Tobsch, V. (2013): "Betreuung von Schulkindern – Ein weiterer Schlüssel zur Aktivierung ungenutzter Arbeitskräftepotenziale?" SOEPpapers on Multidisciplinary Panel Data Research 573, Deutsches Institut für Wirtschaftsforschung (DIW Berlin).

Wagner, G. G., J. R. Frick, and J. Schupp (2007): "The German Socio-Economic Panel Study (SOEP) – Scope, Evolution and Enhancements," *Schmollers Jahrbuch*, 127, 139–169.

Wolf, E. (2002): "Lower Wage Rates for Fewer Hours? A Simultaneous Wage-Hours Model for Germany," *Labour Economics*, 9, 643–663.

——— (2010): "Lohndifferenziale zwischen Vollzeit- und Teilzeitbeschäftigten in Ost- und Westdeutschland," WSI working papers, Wirtschafts- und Sozialwissenschaftliches Institut (WSI), Hans-Böckler-Stiftung.

Wooldridge, J. M. (2010): *Econometric Analysis of Cross Section and Panel Data*, Cambridge, MA: The MIT Press.

Appendix

3A Operationalization of the ADSP Indicator

To construct the indicator for the maternal use of ADSPs for primary school-aged children, I proceed in two steps. In a first step, I use two questions from the household questionnaire to classify primary school-aged children into those making use of an ADSP and those not making use of an ADSP. In a second step, I use the information at the child level to generate an indicator for the use of ADSPs at the maternal level. This second step was necessary because the information at the child level did not map one-to-one to the maternal level for mothers with several primary school-aged children who made use of ADSPs for some but not all of their primary school-aged children.

First step. The first question, which is used to classify children, provides information on the type of school and/or institution that a child attends and is used to identify children who attend primary schools. The second question provides information on the time spent in schools and/or institutions[42] and is exploited to distinguish between primary school students who make use of an ADSP and those who do not make use of an ADSP. Due to a change in the questionnaire, however, there is a small group of children that cannot be unambiguously classified (cf. Marcus et al. 2013, 2016).[43] Yet, Marcus et al. (2013) show that despite this ambiguity official ADSPs use shares for primary school-aged children can be replicated quite well. I classify ambiguous children as users of ADSPs. However, I addressed the ambiguity problem by investigating whether the results are robust to (i) a reclassification of these children as non-users of ADSPs and (ii) to the exclusion of mothers with ambiguous children. I find that the results are robust (results available upon request).

Second step. To aggregate the information from the child level at the maternal level, I use three alternative strategies. First, I classify mothers who make use of an ADSP for at least one of their primary school-aged children as mothers who make use of an ADSP.

[42]Parents had to choose one among the following three answers: mainly in the morning, mainly in the afternoon, and mainly all day.

[43]Prior to 2009, parents were able to provide multiple answers to the first question if, for example, their child attended a primary school and was cared for in an after-school care center. As the second question referred to the time spent in schools *and* other institutions, it is thus unknown for children whose parents indicated both primary school and after-school care center, whether "mainly all day" means that their child attended a half-day school and made use of an after-school care center *or* that their child attended an ADSP and made use of an after-school care center.

Second, I use a more conservative definition and classify mothers who make use of ADSPs for all of their primary school-aged children as mothers who make use of an ADSP. Third, I focus on mothers with one primary school-aged child only and classify mothers who make use of an ADSP for their only primary school-aged child as mothers who make use of an ADSP for primary school-aged children. For the main specification, I decided upon the first strategy, but I show that the results are robust to the two alternative strategies (see section 3.5.4).

I would like to emphasize that it is more meaningful to refer to ADSPs instead of all-day schools in the German context. This is for two reasons: First, in Germany borders between half-day and all-day schools have become more and more blurred over time because many half-day schools have started to provide programs which go beyond the regular school schedules in the morning (Blossfeld et al. 2013). Second, the attendance of an all-day school does not imply the attendance of an ADSP because in Germany roughly 90% of the all-day primary schools make the ADSP available to parents on a voluntary basis (KMK 2014). Therefore, a child who attends an all-day school does not participate in an ADSP, unless the ADSP option is taken by parents. Consequently, it is not the attendance of an all-day primary school but the attendance of an ADSP that should affect maternal labor supply. Given this and the fact that SOEP data measure the attendance of an ADSP rather than the attendance of an all-day primary school, the expression "ADSP" is used in this study.

3B Supplementary Tables and Figures

Table 3B.1. Definition of Variables

Variable name	Definition
Main variables	
Use of ADSP ($ADSP$)	0-1 dummy variable, = 1 if mother makes use of an ADSP for at least one of their primary school-aged children (main specification)
Employed (E)	0-1 dummy variable, = 1 if employed (full-time, part-time, and marginally employed)
Full-time employed (FTE)	0-1 dummy variable, = 1 if full-time employed (only employed mothers)
Instruments	
County received IZBB investments for primary schools	0-1 dummy variable, = 1 if mother lives in a county which received IZBB investments for primary schools (county level, yearly basis)
County never received IZBB investments for primary schools	0-1 dummy variable, = 1 if mother lives in a county which never received IZBB investments for primary schools between 2003 and 2009 (county level)
IZBB investments per primary school/10 000	Year-specific amount of IZBB investments allocated to primary schools in year t per primary school in € 10,000 (county level, yearly basis)
Cumulative IZBB investments per primary school/10 000	Cumulative amount of IZBB investments allocated to primary schools between 2003 and year t per primary school in € 10,000 (county level, yearly basis)
Maternal characteristics	
Age	Age at the time of the interview (in years)
Migration background	0-1 dummy variable, = 1 if any migration background
Married	0-1 dummy variable, = 1 if married
Education	Number of years spent in education
Experience	Number of years of labor market experience
Work preferences	Importance of job success on a 1 to 4 point scale, four categories: missing, low (4 or 3), medium (2), high (1)
Family characteristics	
Number of children	Number of children who live in household and are aged less than 18
Young children present	0-1 dummy variable, = 1 if 0- to 5-year-old children live in the household
Use of other care	Two categories: missing, use of other care (mother makes use of alternative types of care for at least one of their primary school-aged children)
Monthly non-wife net household income/100	Need-weighted deflated monthly net household income net of the mother's labor earnings in € 100
Size of community (area)	Three categories: rural (< 20,000 inhabitants), urbanized (20,000 to 100,000 inhabitants), urban (> 100,000 inhabitants)
Female unemployment rate	Female unemployment rate (county level, yearly basis)
Female part-time employment rate	Female part-time employment rate (county level, yearly basis)

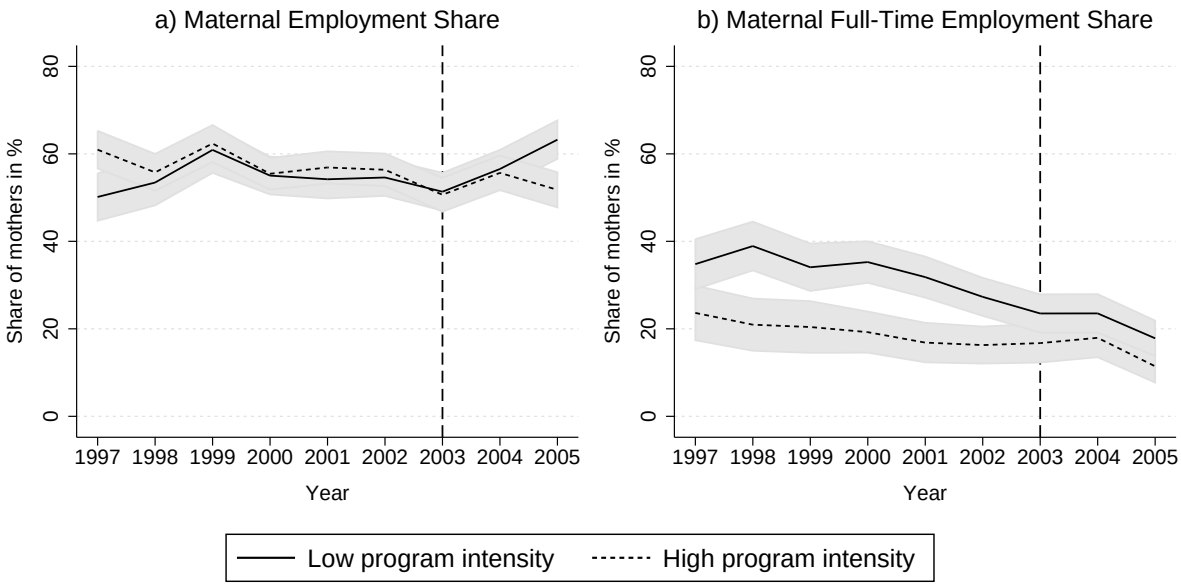

Figure 3B.1. Maternal Employment and Full-Time Employment by IZBB Program Intensity, Germany (1997 – 2005)

Notes: Figure a) shows the share of employed mothers on all mothers with primary school-aged children in Germany who lived in counties with (i) low IZBB program intensity and (ii) high IZBB program intensity, respectively. Figure b) shows the share of full-time employed mothers on all employed mothers with primary school-aged children in Germany again by IZBB program intensity. IZBB program intensity is defined to be low if a county did not receive any IZBB investments or if it received IZBB investments smaller than €36,029.62 per primary school, which corresponds to the median cumulative investment amount per primary school in 2009. The gray shaded areas show 90 percent confidence bands. Estimates of the maternal (full-time) employment share in a given year are based on at least 213 (125) observations, respectively.

Source: SOEPv30 (1997-2005), ReGENESIS (2016) & SPI NRW (2010), own calculations.

Table 3B.2. Inclusion Test

	Extensive margin Employed		Intensive margin[a] Full-time employed	
	Instrument included?			
	No	Yes	No	Yes
Use of ADSPs	0.217***	0.217***	0.361***	0.372***
	(0.078)	(0.079)	(0.116)	(0.117)
Cumulative IZBB investments per primary school/10 000		-0.000		-0.008
		(0.005)		(0.007)
Age (in years)	-0.119***	-0.119***	-0.073***	-0.073***
	(0.009)	(0.009)	(0.015)	(0.015)
Migration background	0.045	0.045	0.497***	0.500***
	(0.077)	(0.077)	(0.120)	(0.121)
Married	0.292***	0.292***	-0.222**	-0.226**
	(0.088)	(0.088)	(0.112)	(0.112)
Education (in years)	0.161***	0.161***	0.139***	0.139***
	(0.015)	(0.015)	(0.022)	(0.022)
Experience (in years)	0.250***	0.250***	0.076***	0.076***
	(0.015)	(0.015)	(0.027)	(0.027)
Experience squared	-0.004***	-0.004***	0.000	0.000
	(0.001)	(0.001)	(0.001)	(0.001)
Number of children	0.002	0.002	-0.110	-0.110
	(0.045)	(0.045)	(0.069)	(0.068)
Young children present	-0.690***	-0.690***	-0.106	-0.110
	(0.063)	(0.063)	(0.108)	(0.107)
Monthly non-wife net household income/100	-0.040***	-0.040***	-0.058***	-0.057***
	(0.006)	(0.006)	(0.010)	(0.010)
Area: urbanized	-0.052	-0.052	0.009	0.011
	(0.084)	(0.084)	(0.115)	(0.115)
Area: urban	-0.071	-0.071	-0.065	-0.048
	(0.095)	(0.095)	(0.149)	(0.148)
Female unemployment rate	-0.298	-0.292		
	(2.076)	(2.089)		
Female part-time employment rate			5.512***	5.388***
			(1.670)	(1.672)
Value of log likelihood	-2,240.57	-2,240.57	-1,212.70	-1,211.86
Observations	5,016	5,016	3,050	3,050

Notes: Probit estimates (coefficients) of maternal labor supply equations. Models additionally include a constant. Standard errors are clustered at the county level and are reported in parentheses. Estimates are robust if linear probability models instead of probit models are estimated.

[a] Probit estimates for the intensive margin are conditional on maternal employment.

** p<0.05 *** p<0.01.

Sources: SOEPv30 (2003-2009), BBSR Bonn (2015), ReGENESIS (2016) & SPI NRW (2010), own estimates.

Table 3B.3. Placebo Test Using Future Investments

	Extensive margin		Intensive margin[a]	
	Use of ADSPs	Employed	Use of ADSPs	Full-time employed
Use of ADSPs		1.116***		-0.338
		(0.239)		(0.469)
Amount of IZBB investments		0.043		-0.111
in year $t+1$/10 000		(0.133)		(0.089)
Cumulative IZBB investments	0.015***		0.020***	
per primary school/10 000	(0.006)		(0.006)	
ρ	-0.525***		0.410	
	(0.134)		(0.261)	
State FE	Yes		Yes	
Year FE	Yes		Yes	
ATE	0.253***		-0.070	
	(0.048)		(0.089)	
Value of log likelihood	-3,655.53		-2,131.21	
Observations	5,016		3,050	

Notes: Bivariate probit estimates (coefficients) of maternal labor supply models. Covariates correspond to those of Table 3.3, except that IZBB investments in year $t+1$ are added to the (full-time) employment equation. Standard errors are clustered at the county level and are reported in parentheses.

[a] Bivariate probit estimates for the intensive margin are conditional on maternal employment.

*** p<0.01.

Sources: SOEPv30 (2003-2009), BBSR Bonn (2015), ReGENESIS (2016) & SPI NRW (2010), own estimates.

Table 3B.4. Placebo Test Using Childless Women

	Extensive margin Employed	Intensive margin[a] Full-time employed
Cumulative IZBB investments per primary school/10 000	0.004	0.002
	(0.003)	(0.003)
Age (in years)	-0.116***	-0.081***
	(0.005)	(0.005)
Migration background	-0.120**	0.014
	(0.060)	(0.073)
Married	0.298***	-0.198***
	(0.051)	(0.060)
Education (in years)	0.119***	0.084***
	(0.010)	(0.010)
Experience (in years)	0.238***	0.132***
	(0.011)	(0.010)
Experience squared	-0.003***	-0.002***
	(0.000)	(0.000)
Monthly non-wife net household income/100	-0.021***	-0.012***
	(0.003)	(0.002)
Area: urbanized	0.065	0.010
	(0.050)	(0.061)
Area: urban	-0.046	-0.028
	(0.046)	(0.088)
Female unemployment rate	-3.160***	
	(0.974)	
Female part-time employment rate		0.018
		(0.908)
State FE	Yes	Yes
Year FE	Yes	Yes
Value of log likelihood	-7,157.79	-5,776.85
Observations	16,214	10,033

Notes: Probit estimates (coefficients) of labor supply equations. Models additionally include a constant. Standard errors are clustered at the county level and are reported in parentheses. Estimates are robust if linear probability models instead of probit models are estimated.

[a] Probit estimates for the intensive margin are conditional on employment.

** p<0.05 *** p<0.01.

Sources: SOEPv30 (2003-2009), BBSR Bonn (2015), ReGENESIS (2016) & SPI NRW (2010), own estimates.

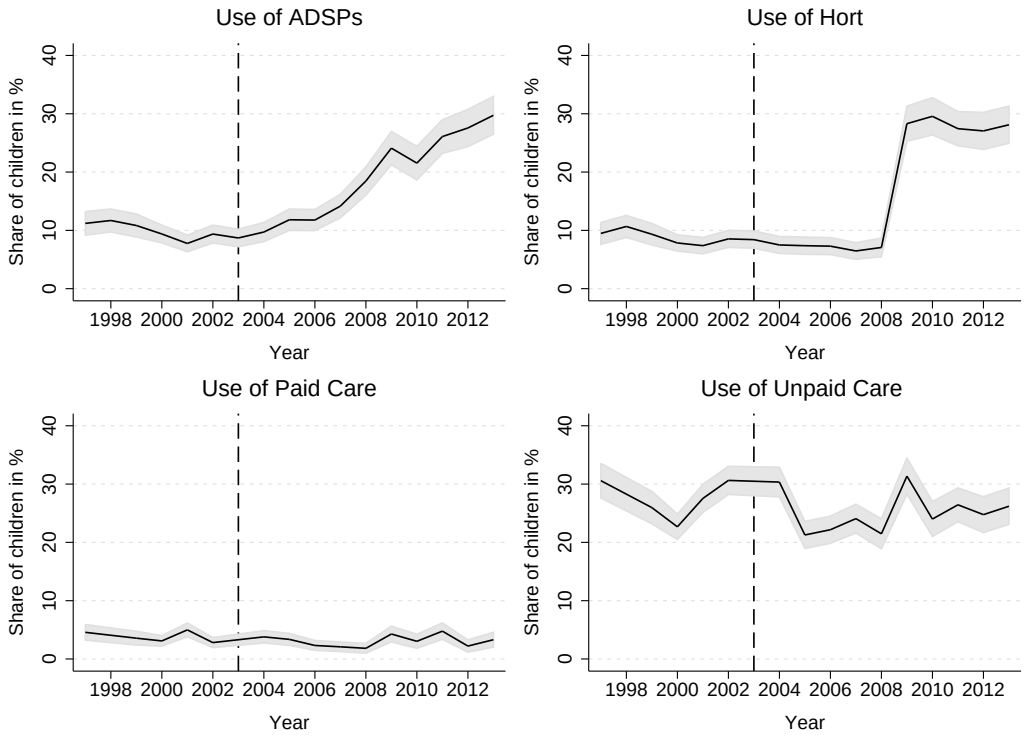

Figure 3B.2. Use of Alternative Types of Care, Germany (1997 – 2013)

Notes: This figure shows the shares of primary school-aged children in Germany who make use of ADSPs, after-school care centers (hort), alternative forms of paid care, and unpaid care, respectively. The gray shaded areas show 90 percent confidence bands. Shares do not add up to 100% as multiple answers were possible and some children were cared for by parents only. The sample encompasses 12,895 child-year-observations. It excludes children of Bavarian and/or self-employed mothers, mothers who were in education and/or did not belong to the working age population, and mothers who had missing information in one of the dependent or explanatory variables. Moreover, it disregards children with missing information in one of the childcare variables. Estimates in a given year rely on at least 532 observations. Information on alternative forms of paid care and unpaid care was not collected in 1998 and 2003.

Source: SOEPv30 (1997-2013), own calculations.

Table 3B.5. 2SLS Estimates of Maternal Labor Supply Models

	Extensive margin		Intensive margin[a]	
	Use of ADSPs	Employed	Use of ADSPs	Full-time employed
Use of ADSPs		0.095		-0.264
		(0.258)		(0.249)
Cumulative IZBB investments per primary school/10 000	0.005***		0.006***	
	(0.001)		(0.002)	
Age (in years)	-0.006***	-0.031***	-0.008**	-0.019***
	(0.002)	(0.003)	(0.003)	(0.004)
Migration background	0.017	0.012	0.013	0.117***
	(0.022)	(0.020)	(0.025)	(0.034)
Married	-0.089***	0.082**	-0.089***	-0.113***
	(0.020)	(0.035)	(0.024)	(0.042)
Education (in years)	0.010***	0.040***	0.013***	0.038***
	(0.003)	(0.005)	(0.004)	(0.007)
Experience (in years)	0.005	0.074***	0.000	0.012**
	(0.003)	(0.004)	(0.005)	(0.006)
Experience squared	-0.000	-0.001***	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Number of children	-0.030***	-0.002	-0.046***	-0.038*
	(0.008)	(0.014)	(0.011)	(0.020)
Young children present	0.008	-0.187***	0.048**	-0.020
	(0.013)	(0.019)	(0.020)	(0.027)
Monthly non-wife net household income/100	-0.003***	-0.010***	-0.003**	-0.011***
	(0.001)	(0.001)	(0.001)	(0.002)
Area: urbanized	0.040***	-0.013	0.026	0.008
	(0.014)	(0.024)	(0.018)	(0.029)
Area: urban	0.126***	-0.013	0.109***	0.038
	(0.027)	(0.043)	(0.035)	(0.048)
Female unemployment rate	-1.697***	-0.112		
	(0.612)	(0.657)		
Female part-time employment rate			0.084	1.269***
			(0.303)	(0.417)
State FE	Yes		Yes	
Year FE	Yes		Yes	
First-stage F-statistic	11.19		13.11	
Observations	5,016		3,050	

Notes: The variable use of ADSPs is instrumented with the cumulative amount of IZBB investments per primary school at the county level. Models additionally include a constant. Standard errors are clustered at the county level and are reported in parentheses.
[a] 2SLS estimates for the intensive margin are conditional on maternal employment.
* p<0.1 ** p<0.05 *** p<0.01.

Sources: SOEPv30 (2003-2009), BBSR Bonn (2015), ReGENESIS (2016) & SPI NRW (2010), own estimates.

Table 3B.6. Robustness Check I: Instrument

	Instrument				
Estimates of	(1) Cumulative IZBB investments per primary school/10 000[a]	(2) Cumulative IZBB investments per primary school/10 000 (exclusion test)[b]	(3) Cumulative IZBB investments per primary school/10 000 (no Thuringia)[b]	(4) Lagged cumulative IZBB investments per primary school/10 000[c]	(5) Amounts of IZBB investments in years t and $t-1$ per primary school/10 000[c,d]
ADSP coefficient in employment equation	1.111***	1.162***	1.016***	1.123***	1.106***
	(0.243)	(0.285)	(0.308)	(0.240)	(0.246)
Coefficient on instrument in ADSP equation	0.015***	0.016***	0.015**		0.028**
	(0.006)	(0.006)	(0.006)		(0.013)
Coefficient on lagged instrument in ADSP equation				0.012**	0.026**
				(0.006)	(0.012)
ρ	-0.521***	-0.540***	-0.460***	-0.527***	-0.518***
	(0.137)	(0.166)	(0.171)	(0.135)	(0.138)
ATE	0.252***	0.261***	0.234***	0.255***	0.251***
	(0.048)	(0.055)	(0.062)	(0.048)	(0.049)
Observations	5,016	4,189	4,816	5,016	5,016

Notes: Bivariate probit estimates for the extensive margin model. Covariates correspond to those of Table 3.3. Standard errors are clustered at the county level and are reported in parentheses.

[a] Model (1) corresponds to the extensive margin model in Table 3.3.

[b] Models (2) and (3) are based on somewhat smaller samples of mothers as I remove mothers who lived in counties with primary schools that received implausible low amounts of IZBB investments (model 2) or mothers who lived in Thuringia (model 3).

[c] Only the (cumulative) amount of IZBB investments allocated to primary schools in a county is lagged by one year. The number of primary schools in a county refers to the current year.

[d] Model (5) uses year-specific IZBB investments rather than cumulative IZBB investments between 2003 and year t.

** p<0.05 *** p<0.01.

Sources: SOEPv30 (2003-2009), BBSR Bonn (2015), ReGENESIS (2016) & SPI NRW (2010), own estimates.

Table 3B.7. Robustness Check II: Alternative Definitions of the ADSP Indicator

	Definition of ADSP indicator		
Estimates of	(1) = 1 if *at least one* primary school-aged child of a mother makes use of an ADSP[a]	(2) = 1 if *all* primary school-aged children of a mother make use of an ADSP	(3) = 1 if the *only* primary school-aged child of a mother makes use of an ADSP[b]
ADSP coefficient in employment equation	1.111***	1.049***	0.876**
	(0.243)	(0.269)	(0.423)
Coefficient on instrument in ADSP equation	0.015***	0.016***	0.015**
	(0.006)	(0.006)	(0.006)
ρ	-0.521***	-0.477***	-0.385
	(0.137)	(0.154)	(0.242)
ATE	0.252***	0.240***	0.204**
	(0.048)	(0.054)	(0.089)
Observations	5,016	5,016	4,249

Notes: Bivariate probit estimates for the extensive margin model. Cumulative IZBB investments per primary school at the county level are used as instrument. Covariates correspond to those of Table 3.3. Standard errors are clustered at the county level and are reported in parentheses.

[a] Model (1) corresponds to the extensive margin model in Table 3.3.

[b] Model (3) is based on a somewhat smaller sample of mothers as I exclude mothers who had more than one primary school-aged child.

** p<0.05 *** p<0.01.

Sources: SOEPv30 (2003-2009), BBSR Bonn (2015), ReGENESIS (2016) & SPI NRW (2010), own estimates.

Table 3B.8. Robustness Check III: Alternative Definitions of the Employment Indicator

	Definition of employment indicator		
Estimates of	(1) Marginal employment = employed[a]	(2) Marginal employment = non-employed	(3) Marginal employment removed from sample[b]
ADSP coefficient in employment equation	1.111***	1.218***	1.014***
	(0.243)	(0.245)	(0.297)
Coefficient on instrument in ADSP equation	0.015***	0.015***	0.016***
	(0.006)	(0.006)	(0.006)
ρ	-0.521***	-0.549***	-0.448***
	(0.137)	(0.138)	(0.167)
ATE	0.252***	0.318***	0.231***
	(0.048)	(0.059)	(0.064)
Observations	5,016	5,016	4,462

Notes: Bivariate probit estimates for the extensive margin model. Cumulative IZBB investments per primary school at the county level are used as instrument. Covariates correspond to those of Table 3.3. Standard errors are clustered at the county level and are reported in parentheses.

[a] Model (1) corresponds to the extensive margin model in Table 3.3.

[b] Model (3) is based on a somewhat smaller sample of mothers as I remove mothers who were marginally employed.

*** p<0.01.

Sources: SOEPv30 (2003-2009), BBSR Bonn (2015), ReGENESIS (2016) & SPI NRW (2010), own estimates.

Table 3B.9. Robustness Check IV: Inclusion of Additional Controls

	Additional sets of controls			
Estimates of	(1) Use of alternative care	(2) Use of alternative care[a]	(3) Work preferences	(4) State-by-year fixed effects
ADSP coefficient in employment equation	1.083***	0.864**	1.031***	1.256***
	(0.236)	(0.430)	(0.282)	(0.249)
Coefficient on instrument in ADSP equation	0.015***	0.015***	0.016***	0.013**
	(0.006)	(0.005)	(0.006)	(0.005)
ρ	-0.535***	-0.416*	-0.482***	-0.610***
	(0.132)	(0.246)	(0.158)	(0.139)
ATE	0.244***	0.192**	0.235***	0.276***
	(0.047)	(0.089)	(0.057)	(0.047)
Observations	5,016	4,201	5,016	5,016

Notes: Bivariate probit estimates for the extensive margin model. Cumulative IZBB investments per primary school at the county level are used as instrument. Covariates correspond to those of Table 3.3, except for Model (4) which uses state-by-year fixed effects instead of state and year fixed effects to control for a possible differential rollout of the IZBB program across states. Standard errors are clustered at the county level and are reported in parentheses.

[a] Model (2) is based on a somewhat smaller sample of mothers as I exclude the year 2003. In 2003, information on alternative types of paid and unpaid care was not collected in SOEP.

* p<0.1 ** p<0.05 *** p<0.01.

Sources: SOEPv30 (2003-2009), BBSR Bonn (2015), ReGENESIS (2016) & SPI NRW (2010), own estimates.

Table 3B.10. Robustness Check V: Endogenous Controls

	Potentially endogenous control	
Estimates of	(1) Non-wife net household income[a]	(2) Female unemployment rate
ADSP coefficient in employment equation	1.132***	1.109***
	(0.307)	(0.239)
Coefficient on instrument in ADSP equation	0.015***	0.014**
	(0.006)	(0.006)
ρ	-0.503***	-0.521***
	(0.182)	(0.135)
ATE	0.262***	0.252***
	(0.061)	(0.048)
Observations	5,015	5,016

Notes: Bivariate probit estimates for the extensive margin model. Cumulative IZBB investments per primary school at the county level are used as instrument. Covariates correspond to those of Table 3.3, except that Model (1) controls for the education of a partner or spouse instead of monthly non-wife net household income and Model (2) excludes the female county-level unemployment rate. Standard errors are clustered at the county level and are reported in parentheses.

[a] Model (1) is based on one observation less due to missing information on the education of the partner or spouse.

** p<0.05 *** p<0.01.

Sources: SOEPv30 (2003-2009), BBSR Bonn (2015), ReGENESIS (2016) & SPI NRW (2010), own estimates.

Chapter 4

Voluntary Pooling of Genetic Risk: A Health Insurance Experiment

Joint with Wanda Mimra and Christian Waibel

Abstract: Scientific and technological advances increasingly allow for better tailoring of health insurance plans to individual health risk profiles. This development questions the sustainability of health plans that feature strong cross-subsidization across different health risk types. An important observation is that the willingness to cross-subsidize in health plans might depend on whether the risk is uncontrollable by individuals, such as genetic risk, or modifiable via health behaviors. In this paper, we provide the results of a laboratory experiment on the willingness to pool genetic risk in health insurance. Subjects' overall health risk has an assigned, uncontrollable genetic risk part and a behavioral risk part, which can be reduced by costly effort. The experimental variation either includes behavioral risk in the pooling of a group insurance scheme or separates it out. Although we observe social preferences for pooling, we observe only a low level of actual genetic risk pooling across the two experimental conditions. This is due to both large heterogeneity in social preferences for pooling across subjects, and the dynamics of the willingness to pay for group insurance in the different experimental markets. Thus, our results indicate that mandatory pooling might be needed if, under the veil of ignorance, a society nevertheless wishes to pool certain forms of heterogeneous risk exposure, such as genetic risk.

JEL classification: I13, C92, D64
Keywords: Health insurance, genetic risk, voluntary pooling, effort, social preferences

Acknowledgements: We wish to thank Pietro Biroli, Ernst Fehr, Steve Heinke, Harald Mayr, Rainer Winkelmann, and seminar participants in Zurich for helpful comments and suggestions. We thank Emina Boudemagh for excellent research assistance.

4.1 Introduction

In the last two decades, scientific and technological advances in detecting, estimating, and monitoring health risks allow for an increased precision of information on individual health risk profiles. The price of sequencing an average human genome has plummeted from about US$10 million in 2007 to a few thousand dollars in the last years. Inexpensive and easily practicable genetic tests are increasingly available for individuals: For example, the US company 23andMe charges people US$99 to see if they have gene variants that put them at higher risk for 120 diseases and whether they carry a known heritable mutation in an additional 50, including cystic fibrosis, sickle cell disease, and Tay-Sachs disease.[1] Most scientists agree that testing for genetic markers is certain to become a far greater part of health care in the future than it is now. On the behavioral side, smart technologies allow us to better track and incentivize health behavior.

These advancements technically allow health insurers to better tailor individual health insurance plans to an individual's particular health risk profile. For the genetic risk part, current legislation in most countries prohibits premium differentiation in health insurance.[2] However, information on health behaviors is increasingly used in pricing: Health insurers in the US and Europe start to provide monetary incentives that are tied to health behaviors that are monitored via, for example, mobile devices. Wearers of devices agree to track their physical activity, such as steps taken, and rewards take the form of credits towards health saving or health reimbursement accounts, lower deductibles, or direct premium discounts.[3] For policymakers, this development poses several questions: Should the pricing of health insurance plans, in both public and private markets, be generally allowed to condition on tracked health behaviors? Should the general prohibition of using genetic information in health insurance be upheld? Regarding the latter, increasing availability of inexpensive genetic tests could lead to a call for optional revelation of genetic risk

[1] Costs of genetic testing for individuals are strongly influenced by the market structure and patent protecting in particular countries. For instance, the cost of BRCA testing ranges from US$475 to about US$4,000.

[2] In the US, the 2008 Genetic Information Nondiscrimination Act (GINA) prohibits discrimination by health insurance plans based on an individual's genetic information. The GINA does not extend to life insurance, however.

[3] The US health insurance provider UnitedHealth, for example, offers policyholders on the Motion F.I.T. program up to US$4 per day in credits applied towards their health saving or reimbursement account if they use a Fitbit to track their physical activity and reach one or more fitness oriented goals. Premium discounts of up to 15 percent are granted by another major US health insurance provider upon tracking exercising progress and reaching activity goals.

information by policyholders. Theoretically, this could trigger information unraveling in the sense of Milgrom (1981), i.e. the best (lowest genetic risk) types voluntarily disclose their genetic information to receive a lower premium or other improved terms, and then the next best types have an incentive to do so as well, until there is full disclosure in the market. Now, a crucial observation is that this form of unraveling with pure hidden information is actually efficient, and even allows for additional efficiency increases resulting from a better adjustment of health behaviors. However, it entails severe distributional consequences, and further punishes individuals that were already unlucky in the genetic lottery. This is one of the motivations behind the current legislation of prohibiting the use of genetic information in health insurance plans.

In this paper, we investigate whether there is support for this view in an incentivized health insurance experiment: We analyze the willingness to pool genetic risk in health insurance when a fully individually risk-adjusted health insurance plan is also available. Thus, in its simple form we test for social preferences in the context of health insurance. Our experiment further makes use of the observation that better health risk detection and monitoring allow to separate health risks that are uncontrollable by an individual, such as the genetic predisposition, from health risks that stem from an individual's health behavior. In particular, with our experimental variation, we test whether more mutualization of genetic risk can be achieved by separating and individually pricing risk components that are within the control of an individual. Separating out risk components that are the result of an individual's effort decisions eliminates free-riding incentives and might thereby induce more people to be willing to mutualize exogenous heterogeneity in risk exposure, here genetic risk differences.

To do so, in our experiment subjects face the risk of illness, which is comprised of two parts: a genetic risk component and a behavioral risk component. Subjects are exogenously assigned an either high or low genetic risk and they decide on a costly preventive effort, which reduces their behavioral risk. To make the health context salient, by choosing a higher effort level, subjects do not only decrease their risk of illness and thereby reduce their health insurance premiums in the experiment, but they also increase the probability of winning a voucher for a preventative health screening at the local sports facilities. Subjects state their willingness to pay (WTP) for a group insurance scheme that pools health risks, and decide on individual risk-based insurance or no insurance for

the case that they might not be included in the group insurance.[4] The outside option, individual insurance, is fully individually risk-adjusted, i.e. it prices individual health risk including behavioral risk with an actuarially fair premium.[5] In the experiment, we vary the extent of risk pooling in the group health insurance between full pooling (FP), in which the full health risk of group insurance participants is pooled, and genetic pooling only (GPO), in which the genetic risk component is pooled, but participants in group insurance receive individual premium discounts based on their preventive effort.

We find that across experimental conditions, about half of the subjects who were assigned a low genetic risk, i.e. the types who would need to cross-subsidize high genetic risk types in a group insurance, exhibit social preferences that manifest in a WTP for group insurance that exceeds their individual insurance premium. In the GPO condition, with pooling of genetic risk only, low genetic risk subjects were roughly ten percentage points more likely to indicate a WTP that exceeded their individual insurance premium than in the FP condition, with this difference being particularly large in the first five periods. However, due to both large heterogeneity in social preferences in terms of the willingness to pool health risks across subjects, and the dynamics of the WTP for group insurance in the different experimental markets, we observe only a low level of voluntary genetic risk pooling in both experimental conditions. These results highlight the difficulty of achieving pooling of heterogeneous risks in health insurance when private markets with fully risk-adjusted premiums are available. Thus, our results indicate that mandatory pooling might be needed if, under the veil of ignorance, a society nevertheless wishes to pool certain forms of heterogeneous risk exposure, such as genetic risk.

Related Literature

Our health insurance experiment is related to three strands of literature: The (experimental) literature on preferences for redistribution, the literature on the role of heterogeneous endowments in public goods games, and the new experimental literature on health insur-

[4]For simplicity, we refer to the individual risk-based insurance as individual insurance in what follows. The idea behind this is that an insurer still ensures a pool of individuals, but charges each individual her actuarially fair premium. This is in contrast to group insurance, in which the premium for an individual depends not only on her own risk profile but also on that of the group. The group insurance premium is endogenously determined ex post. Subjects who indicate a WTP that is at least as high as the group insurance premium that would result if they were included in group insurance, participate in group insurance.

[5]Thus, here, we are not interested in the standard ex-ante moral hazard problem in insurance, but only in the moral hazard problem of free-riding under pooling in a group insurance scheme.

ance demand. Our paper combines the question of fairness and inequality views with that of incentive problems present in public goods games in the context of health insurance choice.

The literature on preferences for redistribution suggests that people are averse to deviations from both an equal income distribution (e.g., Fehr and Schmidt 1999, Bolton and Ockenfels 2000, Engelmann and Strobel 2004) and an income distribution that is proportional to work effort (e.g., Konow 2000, Frohlich et al. 2004, Cappelen et al. 2007, 2010, 2013).[6] In the context of risky situations, Cappelen et al. (2013), for instance, analyze fairness views in an experiment in which subjects ex ante face the same choices between a risky and a safe alternative, and there is a redistribution choice for ex-post income. The authors find that most participants favor not equalizing ex-post inequalities that result from different choices, but that the converse holds true for ex-post inequalities resulting from differences in luck among risk-takers. Mollerstrom et al. (2015) conduct a laboratory experiment in which spectators redistribute ex-post resources between two agents facing a situation in which part of the outcome is controllable, whereas another part is not. They find that many spectators condition their allocation for bad uncontrollable luck on an agent's decision on controllable luck exposure, even though the two types of luck are independent. Our experiment differs from these and related works in three important ways: First, we consider ex-ante heterogeneity in risk exposure and social preferences with respect to it. Second, contrary to including an explicit ex-post redistribution stage as in the previous literature, which makes redistribution particularly salient, redistribution in our experiment is implicit in the choice of health insurance schemes. Third, in our experimental variation, we compare two health insurance systems that by design either include or fully exclude controllable risk exposure in its redistributive scheme. Moreover, we also compare the health insurance and redistribution choices in the incentivized experiment with a subject's preferences for distinct health insurance systems expressed in a post-experimental survey.

The role of heterogeneity in and origin of endowments for contributions in public goods games is analyzed in Cherry et al. (2005), Oxoby and Spraggon (2013), and Kingsley (2016). Cherry et al. (2005) show that individuals provide less to the public good in a

[6]However, Ku and Salmon (2013), varying the source of initial inequality between random, meritocratic, and rewarding uncooperative behavior, find that random assignment leads to the most tolerance for disadvantageous inequality.

group with heterogeneous than homogeneous endowments. The authors find no evidence that the origin of endowment influences the level of contributions, however, thereby challenging the hypothesis that positive contributions in a public goods game are an artifact of endowment origin. Oxoby and Spraggon (2013) show that heterogeneous origins of endowment may lead to a lower public goods provision if minorities exist. The authors argue that the lack of identification among minorities causes the decrease in contributions. Kingsley (2016) studies the effectiveness of punishment under heterogeneous and homogeneous endowments. Whereas contributions increase when introducing punishment under homogeneous endowments, contributions do not change when endowments are heterogeneous. In contrast to the public goods literature, our paper focuses on a health care setting with ex-ante homogeneous endowments but heterogeneous risk exposure. Hence, our experiment mirrors a situation with ex-post instead of ex-ante heterogeneously distributed wealth.

In the experimental health insurance literature, Buckley et al. (2012) analyze experimentally how characteristics of the public health system affect a subject's WTP for parallel private health insurance. Buckley et al. (2012) find that average WTP is lower when the public system allocates health care based on need rather than randomly. Closest to our paper in experimental set-up is Gajdos et al. (2017). They consider an experimental health insurance game in which subjects differ with respect to wealth, health risk profile, and observable effort choice. Subjects provide their WTP for a mutual insurance in which the overall health risk of members of this insurance is pooled. In a within-subject design, Gajdos et al. (2017) consider the effect of an informational boost after some periods, in the form of both a contribution simulator to see how mutual insurance works and how a subject's contribution was shared, but they also investigate the effect of a health insurance framing relative to an originally neutral framing. Gajdos et al. (2017) find that the informational boost temporarily increases the WTP for mutual health insurance. Due to multiple simultaneous variations, the causes of this effect cannot be disentangled, however.

The remainder of this paper is organized as follows: Section 4.2 provides information on the experimental set-up including parametrization and theoretical predictions. Section 4.3 presents the results. Section 4.4 discusses some implications of the results and concludes.

4.2 Experiment

We start by discussing the experimental set-up. Then, we present a simple theoretical framework to illustrate the role of social preferences in the health insurance context. Based on this framework, we derive our main testable hypothesis.

4.2.1 Experimental Design

In our experiment, we apply a between-subject design to vary the degree of risk pooling between *full pooling* and *genetic pooling only*. Subjects are randomly assigned to one of the two conditions.[7] Matching groups of eight subjects, which are called societies, are implemented in both conditions. The assignment to a matching group is random and does not change throughout the experiment. Within a matching group, four subjects are randomly assigned a low genetic risk, while the other four subjects are assigned a high genetic risk. The distribution of genetic risk types within a matching group is common knowledge whereas subject's individual risk type is private information. In both conditions, there are six matching groups.

The experiment involves ten periods. In each period and in both conditions, subjects have an initial endowment of 1000 ECU. Subjects face the risk to turn ill. Illness requires costly treatment of 700 ECU. The overall probability to turn ill is given by the sum of the genetic risk and a behavioral risk component. Whereas genetic risk is non-modifiable, behavioral risk depends on a subject's preventive effort. Low genetic risk types have a genetic risk of 20% to turn ill, high genetic risk types of 40%. The initial behavioral risk amounts to 20% for both genetic risk types, such that the overall probability to turn ill before preventive effort is 40% for a subject that is assigned the low genetic risk, and 60% for a subject that is assigned the high genetic risk.

In each period and in both conditions, subjects make the same three decisions. First, each subject chooses a level of preventive effort ranging from zero to ten. Preventive effort linearly reduces a subject's behavioral risk but is costly. Effort costs are convex in the level of preventive effort. A unique feature of our experimental design is that the subject's effort choice is tied to the probability of winning a voucher for a health preventative

[7] Note that we refer to conditions instead of treatments to distinguish between the treatment for subjects that turned ill in a given period and experimental conditions.

Table 4.1. Overview of Preventive Effort Parameters

Level of preventive effort	Reduction in behavioral risk (in percentage points)	Probability to win the voucher (in percent)	Costs for health prevention (in ECU)
0	0	0	0
1	2	1	8
2	4	2.5	18
3	6	4.5	30
4	8	7	46
5	10	10	66
6	12	13.5	90
7	14	17.5	118
8	16	22	150
9	18	27	186
10	20	33	226

measure that aims at detecting inefficient and harmful movement patterns.[8] The voucher has a monetary value of US$65 and entitles the winner to take the preventative measure free of charge. We use the voucher to make the health prevention decision more salient and to increase heterogeneity in subjects' effort choices. Table 4.1 provides an overview of the range of effort levels and the corresponding reductions in behavioral risk, likelihoods of winning the voucher, and costs of providing preventive effort.

Second, subjects make their health insurance choice. This choice involves two simultaneous decisions: Each subject states her WTP for group insurance. Subjects who indicate a WTP that is at least as high as their group insurance premium are group insured.[9] Moreover, each subject decides whether to purchase individual insurance at her actuarially fair premium or to stay uninsured for the case that she might not be included in the group insurance. Both types of health insurances provide full coverage for individuals.[10]

Insurance premiums for the group insurance are calculated such that the group insurance makes zero profits in expectation, i.e. the sum of the group insurance premiums

[8]More specifically, the health prevention voucher is for a "Functional Movement Screen". The Functional Movement Screen is a test, which was developed in the Unites States. It is used to detect weaknesses in movement orders and to improve the course of motion in order to prevent degeneration and damage of the musculoskeletal system. In the long run, degeneration as well as damage of the musculoskeletal system causes strong pain and may lead to high treatment costs (e.g., due to the treatment by an orthopedic specialist or a physiotherapist).

[9]The group insurance premium that is relevant for comparison with a subject's WTP is the one that would result if the subject was included in the group insurance. The algorithm used to compute the group insurance premium maximizes (i) the number of subjects participating in the group insurance, (ii) the number of subjects with high genetic risk in group insurance, and then randomly selects among the remaining possible group insurance candidates.

[10]Figures 4A.1 and 4A.2 in Appendix 4A show the decision screens for the preventive effort decision and the health insurance decisions.

Table 4.2. Overview of Experimental Conditions

	FP condition	GPO condition
Preventive effort	Yes	Yes
Premium group insurance	pooling of genetic + behavioral risk	pooling of genetic risk + indiv. priced behavioral risk
Premium individual risk-based insurance	individually priced genetic + behavioral risk	individually priced genetic + behavioral risk
Outside options	individual risk-based insurance, no insurance	

exactly covers the expected treatment costs of the subjects in a matching group that are insured by group insurance. The crucial difference between our two experimental conditions is the degree of risk pooling in group insurance. In the FP condition, genetic risk *and* behavioral risk post effort are pooled, whereas in the GPO condition, only genetic risk is pooled while behavioral risk is individually priced. That is, in the FP condition, all members of the group insurance pay the same premium. This premium is based on the average overall risk of illness of group insurance members and it takes preventive effort choices of all group insurance members into account. In the GPO condition, the premium consists of two parts: The first part is identical for all group insurance members and is based on their average genetic risk of illness. The second part is individually priced and is based on the remaining behavioral risk after exerting preventive effort. It appears as an individual premium discount for subjects, which depends on their individual preventive effort decision. Table 4.2 summarizes the above outlined decision sequence and highlights the differences between the two conditions.

In each period and in both conditions, subjects' profits depend on the level of preventive effort, the insurance status, and, if they are not insured, on the state of illness. Insured subjects receive the initial endowment and pay their insurance premium as well as their effort cost for health prevention. Non-insured subjects receive the initial endowment and pay the cost for prevention. In addition, non-insured subjects pay the treatment cost of 700 ECU if they turn ill. To avoid income effects, at the end of the experiment one of the ten periods is randomly selected to be payoff-relevant.

At the end of each period, subjects observe a summary screen of the current period (see Figure 4A.3 in Appendix 4A). This summary screen also provides information on group insurance: existence of group insurance, number of members, number of members

with high genetic risk, and premium. This feedback allows subjects to learn about the other subjects' social preferences for pooling of health risk in their matching group over time.

4.2.2 Experimental Procedure

The experimental sessions were conducted in October and November 2017 at the ETH Decision Science Laboratory. 96 subjects participated in the experiment, 48 in each condition. Participants were, on average, 22 years old. 53.1% of the participants were female. All participants were enrolled students. More than one third of the participants were enrolled for natural sciences, roughly one fifth for engineering, 7.2% for medicine, 6.2% for humanities, and 13.5% for economics. The remaining 15.6% of participants were enrolled in other subjects.

We performed the experiment using z-Tree (Fischbacher 2007). Subjects participated in exactly one session. The average time per session was about two hours. Participants earned 50 CHF, on average. A comprehensive set of control questions ensured that all participants understood the sequence of decisions in the experiment and the payoff consequences.

After the main experiment, we elicited risk preferences using the Holt-Laury task (Holt and Laury 2002) and altruism using the dictator game (e.g., Kahneman et al. 1986, Engel 2011). One of the two games was randomly selected to be payoff-relevant. We also launched a post-experimental questionnaire to collect further information on subjects. We used items of the Falk preference module to obtain additional information on risk aversion, altruism, and reciprocity (Falk et al. 2016). Moreover, we collected information on demographics (e.g., age, gender, major). Importantly, we also asked subjects how they would vote on four different health insurance systems that differed with respect to the degree of risk pooling, whether they were using a health app, and whether they would be willing to share information about their health with their health insurance provider. The information on voting preferences will allow us to compare stated with revealed preferences.[11] Figure 4.1 summarizes the timeline of the experiment.

[11] See Appendix 4B for the exact wording of the survey questions and answer possibilities.

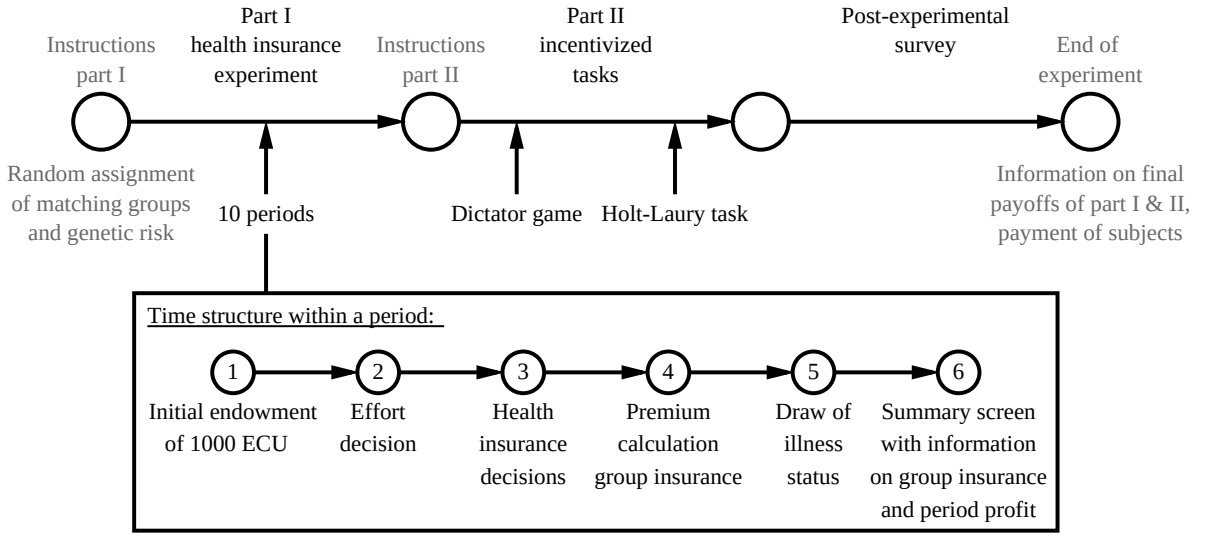

FIGURE 4.1. EXPERIMENTAL TIMELINE

4.2.3 Framework and Predictions

Consider a society of N individuals. Each individual has wealth $y > 0$ and faces the risk of illness. Individuals differ with respect to their genetic predisposition to turn ill, i.e. they may be either a high genetic risk type (H) or a low genetic risk type (L). The overall probability of illness, p_θ, for an individual of type θ, $\theta \in \{H, L\}$, depends on both the genetic risk π_θ, where $0 < \pi_L < \pi_H < 1$, and a preventive effort decision e in the following way:

$$p_\theta(e) = \underbrace{\pi_\theta}_{\substack{\text{genetic risk} \\ \text{component}}} + \underbrace{z - h(e)}_{\substack{\text{behavioral risk} \\ \text{component}}} . \tag{4.1}$$

The genetic risk component, π_θ, is non-modifiable, but the behavioral risk component, $z - h(e)$, can be reduced by costly preventive effort $e \in [0, 1]$. As in our experiment, we assume that effort reduces the behavioral health risk linearly and type-independently, i.e. $h(e) = a \cdot e$, $a \in (0, z]$.[12] Effort costs, $c(e) > 0$, are increasing and convex.

Individuals who turn ill incur monetary costs M, where $0 < M < y$. To ensure against this loss, they can purchase health insurance. Two different forms of health insurance are available, an insurance that we will call, for simplicity, individual insurance, and a group insurance. Both insurances provide full coverage, but they differ with respect to

[12]We only consider the simple case in which the marginal benefit of effort in terms of a reduction of the probability of turning ill is the same across types. This may well be different. For instance, the marginal benefit of effort might be higher for H-types, such that effort decreases the difference in the overall probabilities of turning ill across types.

their premium: The individual health insurance prices both components of health risk according to the actuarially fair rate for an individual, such that the total premium of a θ-type individual with preventive effort e for the individual insurance is $P_\theta^I(e) = (\pi_\theta + z - a \cdot e) M$. That is, the underlying assumptions are that effort is observable to insurance providers, and that effort information is used to charge a premium that fully internalizes the risk reduction resulting from preventive effort. Group insurance pools its members either on the genetic risk component (GPO condition) or on overall risk, i.e. both the genetic and the behavioral risk components (FP condition). Assuming that the group insurance is overall making zero expected profits, the group insurance premiums for a θ-type individual with effort e are given by

$$\begin{aligned} P_\theta^G(e) &= (\bar{\pi} + z - a \cdot e) M \qquad \text{(GPO condition)} \\ P_\theta^G(e) &= (\bar{\pi} + z - a \cdot \bar{e}) M \qquad \text{(FP condition)}, \end{aligned}$$

where $\bar{\pi}$ denotes the average genetic risk of group insurance members and $\bar{e}$ denotes the average effort of group insurance members. Thus, under GPO, the premium of an individual fully internalizes the individual's preventive effort whereas, under FP, risk reduction from preventive effort is pooled across the group insurance members, such that there are free-riding incentives.[13]

In this paper, our focus is on whether individuals voluntarily select group insurance. Given the outside option of the individual insurance with an actuarially fair premium that is fully risk-adjusted and fully internalizes preventive effort, it is easy to see that if an individual's utility only takes into account her own payoff, L-type individuals who have a choice between individual and group insurance are not willing to pool with H-types in group insurance.

To model social preferences in this context, we assume that agents are inequity averse with respect to genetically caused income differences, using a variant of the Fehr-Schmidt-model (Fehr and Schmidt 1999). To highlight the role of inequity aversion with respect to genetically caused income differences, we consider the benchmark case of a society of $N = 2$ individuals, where one individual is an H-type and the other individual is a L-type.

[13]Note that the situation considered here is different from the standard moral hazard problem under insurance where the insurance premium cannot be based on effort due to unobservability of effort.

For simplicity, we also assume that utility is linear in consumption.[14] Therefore, in the case of a society with two individuals, the utility function for individual i is given by

$$\begin{aligned} U_i\left(\Pi_i^k, \Pi_{-i}^{gen,l}, \Pi_i^{gen,k}\right) = \Pi_i^k &- \alpha_i \max\left(\Pi_{-i}^{gen,l} - \Pi_i^{gen,k}, 0\right) \\ &- \beta_i \max\left(\Pi_i^{gen,k} - \Pi_{-i}^{gen,l}, 0\right), \end{aligned} \tag{4.2}$$

where $\Pi_i^k = y - P_i^k - c(e_i)$ is the consumption of individual i given her insurance choice k, $\Pi_{-i}^{gen,l}$ is the genetic income component of the other individual $-i$ given the other individual's insurance choice l, and $\Pi_i^{gen,k}$ is the genetic income component of individual i given her insurance choice k. In equation (4.2), the second (third) term measures the loss from disadvantageous (advantageous) income inequality that stems from differences in genetic risk exposure. Under the standard assumption, $\alpha_i \geq \beta_i \geq 0$, advantageous inequality is not more important than disadvantageous inequality. The crucial difference between group insurance and individual insurance comes into play with respect to these two inequity terms: If both individuals have group insurance, either of GPO or FP form, both inequity terms are equal to zero, since the insurance premium part for the genetic risk component is the same across both individuals, such that genetically determined income differences are equalized. Under individual insurance, however, the second term is non-zero for the H-type individual and the third term is non-zero for the L-type individual.[15]

We analyze the case in which the individuals make the following two decisions: They decide on whether to purchase individual or group insurance, and they decide on their preventive effort to reduce their risk of illness.[16] A group insurance that pools the individuals on genetic risks will exist if both individuals prefer the group insurance to the individual insurance. This will be the case if for both individuals the utility with group insurance is at least as high as the utility with individual insurance.

Now, first consider optimal effort choices, given that an individual is insured under either group or individual insurance. Under individual insurance, both individuals will choose the same optimal effort level, e^*, which is determined by $c'(e^*) = a \cdot M$, since effort is fully internalized in the premium adjustment. Under group insurance, the optimal

[14] This assumption is only made for ease of exposition and it does not affect the results on the role of inequity aversion.

[15] See Appendix 4C for the utility functions of the H-type and L-type individuals under group and individual insurance.

[16] Under the assumption of linear consumption utility, individuals receive the same utility from individual insurance as without insurance, such that we discard the no insurance option here.

effort choice depends on whether there is group insurance of the GPO or the FP form. In the GPO case, effort incentives under group insurance are identical to those under individual insurance. Hence, both individuals provide effort e^*. In the FP case, effort incentives under group insurance differ from those under individual insurance because of the free-riding incentives that exist under pooling of behavioral risk. In this case, if both individuals are insured with group insurance of the FP form, both individuals will choose $e^o < e^*$, which is determined by $c'(e^o) = \frac{1}{2} \cdot a \cdot M$. Marginal benefits of effort are reduced by the factor $\frac{1}{2}$ because, under FP, benefits of effort are split equally among both group insurance members.

Next, consider optimal insurance choices. We start with the *benchmark case of no social preferences*, i.e. $\alpha_i = \beta_i = 0$ for both individuals. In this case, as indicated above, a group insurance exists neither under GPO nor FP because the L-type individual is never willing to pool genetic risk. This is because her utility with group insurance is strictly lower than her utility with individual insurance. Thus, without social preferences both individuals purchase individual insurance and provide the optimal effort level e^*. Now, consider the *case with social preferences*. Again, the crucial individual is the L-type individual, since this individual needs to cross-subsidize the H-type individual under group insurance. Thus, the predictions depend on the advantageous inequity aversion parameter of the L-type individual, β_L, and differ between the two experimental conditions GPO and FP. The L-type individual is willing to select group insurance over individual insurance if she is sufficiently inequity averse, i.e. if $\beta_L \geq \underline{\beta}$. In the GPO condition, $\underline{\beta} = \frac{1}{2}$ and, in the FP condition, $\underline{\beta} = \frac{1}{2} + \kappa\left(e^*, e^o\right)$ with $\kappa\left(e^*, e^o\right) = \frac{(ae^*M - c(e^*)) - (ae^oM - c(e^o))}{(p_h - p_l)M} > 0$. That is, in the FP condition the L-type individual must have a higher advantageous inequity aversion parameter, β_L, than in the GPO condition to select group insurance over individual insurance. This effect is due to the efficiency loss from free-riding under FP group insurance, which lowers the utility under FP group insurance.

This simple analysis, based on a society with one low and one high genetic risk type individual, shows that we should only observe group insurance if subjects with low assigned genetic risk of illness are sufficiently inequity averse. In our experiment, a WTP for group insurance that exceeds the insurance premium for individual insurance, both under GPO and FP, indicates the presence of social preferences as the outside option of individual insurance is always available. In Section 4.3, we will therefore start with an

analysis of the WTP in comparison to the individual insurance premium, with a focus on the crucial L-type individuals. The above analysis of the difference in threshold values for β_L between the two experimental conditions also motivates our main hypothesis for differences across the two conditions:

Hypothesis. *Under GPO, there is on average more group insurance than under FP.*

4.3 Results

We begin by giving a brief overview of the results before investigating in depth (1) how social preferences manifest at the individual level in the health insurance context and (2) how this translates to voluntary mutualization of heterogeneous health risks at the societal level. Then, we analyze whether there is a discrepancy between incentivized choices in the experiment and non-incentivized voting decisions elicited in the post-experimental survey.

4.3.1 Overview of Results

Table 4.3 summarizes insurance and preventive effort choices, and resulting insurance outcomes at the societal level. On average, subjects are willing to pay 250 ECU for group insurance. To put this in context, the range of insurance premiums under individual insurance is between 140 ECU (highest effort level) and 280 ECU (lowest effort level) for low genetic risk types and correspondingly between 280 ECU and 420 ECU for high genetic risk types. In both conditions, subjects with high genetic risk have, on average, a higher WTP for group insurance than subjects with low genetic risk (p-value = 0.0277, Wilcoxon signed-rank test), which is expected given their higher overall risk. Moreover, for subjects with low genetic risk, average WTP is higher in the GPO than the FP condition. The difference is not statistically significant, however. Average effort amounts to 4.9, and neither differs substantially across genetic risk types nor experimental conditions.

Turning to the resulting health insurance outcomes, a group insurance is 1.6 times more likely to exist in the GPO condition, when compared to the FP condition. The share of subjects participating in group insurance is also higher in the GPO condition. Differences in participation are particularly striking for subjects with high genetic risk: In the GPO condition, almost every second subject with high genetic risk participates in

Table 4.3. Choices and Insurance Outcomes, Total and by Genetic Risk Type and Experimental Condition

		Low risk types		High risk types	
	Total	FP condition	GPO condition	FP condition	GPO condition
WTP for group insurance (in ECU)	250	163.3	175.3	332.3	329.3
Existence of group insurance (in %)	69.2	53.3	85.0	53.3	85.0
Participation in group insurance (in %)	23.1	12.9	16.7	17.9	45.0
Preventive effort	4.9	4.9	4.6	4.9	5.2
Observations	960	240	240	240	240

Notes: Subject-period-observations (e.g., 960 = 96 subjects × 10 periods). 20 ECU = 1 CHF. Possible group insurance premiums range between 140 ECU (low risk subjects with effort of ten only) and 420 ECU (high risk subjects with effort of zero only).

a group insurance whereas, in the FP condition, less than one subject with high genetic risk per society participates, on average. The difference in participation shares across experimental conditions reflects both a higher propensity of existence of a group insurance (extensive margin) and a larger size of group insurance conditional on existence (intensive margin) in the GPO condition. In the GPO (FP) condition, 2.8 (2.2) subjects participate in a group insurance, on average. Among them, roughly 2 (1.4) have a high assigned genetic risk of illness.

In this paper, we are particularly interested in whether pooling of heterogeneous genetic risk types can be achieved when people have the choice between mutualization and fully risk-adjusted insurance premiums. The descriptive results presented before do not distinguish between group insurance that pools different genetic risk types and group insurance that does not, however. In Figure 4.2, we therefore explicitly distinguish between mixed (i.e., a group insurance that pools different genetic risk types) and non-mixed (i.e., a group insurance which does not pool different genetic risk types) group insurance when presenting insurance shares by genetic risk type and experimental condition. We observe that patterns for participation in mixed and non-mixed group insurance are similar to those previously described for group insurance overall. Mixed group insurance exists in both experimental conditions, confirming that some subjects are willing to cross-subsidize other participants in group insurance to equalize income differences that stem from differences in uncontrollable risk exposure, independent of the exact mutualization scheme. However, under GPO, we observe a higher participation of low genetic risk types, as well as an overproportional increase in the share of cross-subsidized high genetic risk types.

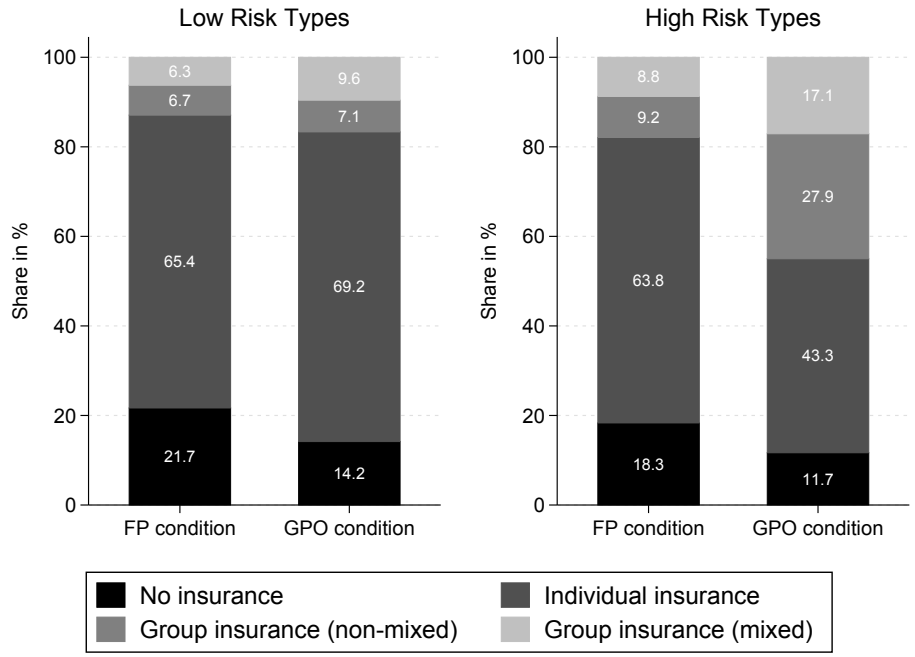

Figure 4.2. Insurance Shares by Genetic Risk Type and Experimental Condition

Yet, Figure 4.2 also shows that the large majority of subjects ends up with individual or no insurance if participation in group insurance is voluntary.

4.3.2 The Willingness to Pool Genetic Risk

To better understand our overall results on mutualization, we first analyze the manifestation of social preferences at the individual level. Social preferences for redistribution in our health insurance set-up translate to a willingness to cross-subsidize. We use the net WTP, which is given by the difference between a subject's WTP for group insurance and her premium under individual insurance, to measure social preferences. More specifically, a positive net WTP of subjects with low genetic risk indicates the willingness to cross-subsidize high genetic risk types.[17] Figure 4.3 shows the share of subjects with positive net WTP pooled across societies and periods by genetic risk type and experimental condition. Overall, the share of subjects indicating a positive net WTP is with 20% considerable.[18] Again, we observe differences across experimental conditions that are in line with the theoretical predictions: In the FP condition, roughly 15% of the WTP responses are such that they exceed a subject's individual insurance premium. In the GPO

[17]Subjects with a positive net WTP must exhibit social preferences because they are willing to sacrifice some fraction of their income to reduce premium differences within their society.

[18]On average, subjects with positive net WTP are willing to give up 95.80 ECU.

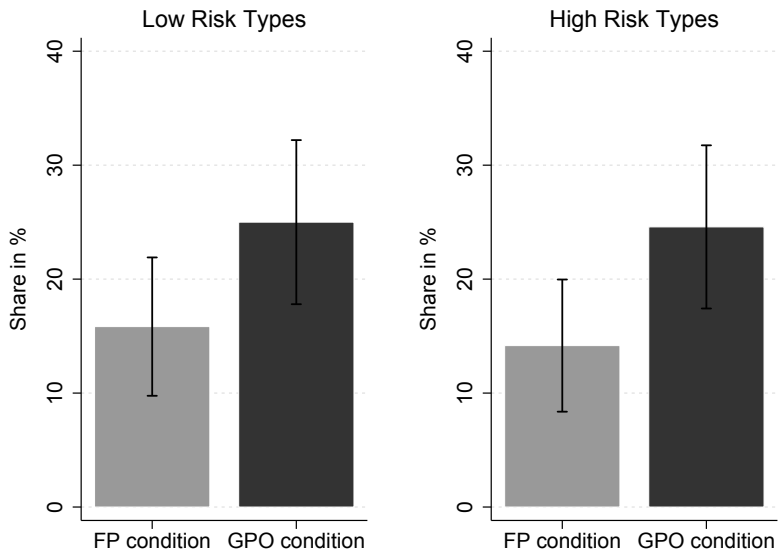

Figure 4.3. Share of Subjects with Positive Net WTP by Genetic Risk Type and Experimental Condition

Notes: The net WTP equals the WTP for group insurance minus the premium for individual insurance.

condition, this share is at 24.9% roughly 1.7 times larger. This difference in positive net WTP shares across experimental conditions is statistically significant (p-value = 0.0247, Mann-Whitney U-test). In both experimental conditions, positive net WTP shares do not differ between subjects with low and high genetic risk of illness. This finding is curious, since high genetic risk types, given their effort level, cannot increase their utility (including inequity aversion for genetically caused income differences) with a positive net WTP. One explanation for this finding is that by stating a high WTP, leading to a positive net WTP, subjects with high genetic risk nevertheless express their preferences for risk pooling.

Figure 4.4 depicts the share of subjects with positive net WTP over time, again by genetic risk type and experimental condition. It shows for both genetic risk types that the difference across experimental conditions also holds over time: In all but two periods, the share of subjects with positive net WTP is larger in the GPO than in the FP condition, although for low genetic risk types the difference between the two experimental conditions decreases over time.[19]

[19]Note that this decline should be less pronounced if the size of a society increases or the share of subjects with high genetic risk decreases because it becomes more likely that several low genetic risk types cross-subsidize one high genetic risk type, thereby decreasing the burden for each low genetic risk type. The lower burden associated with cross-subsidization may then motivate additional low genetic risk subjects to participate in group insurance. Both effects are expected to be stronger in the absence of free-riding incentives, i.e. in the GPO condition.

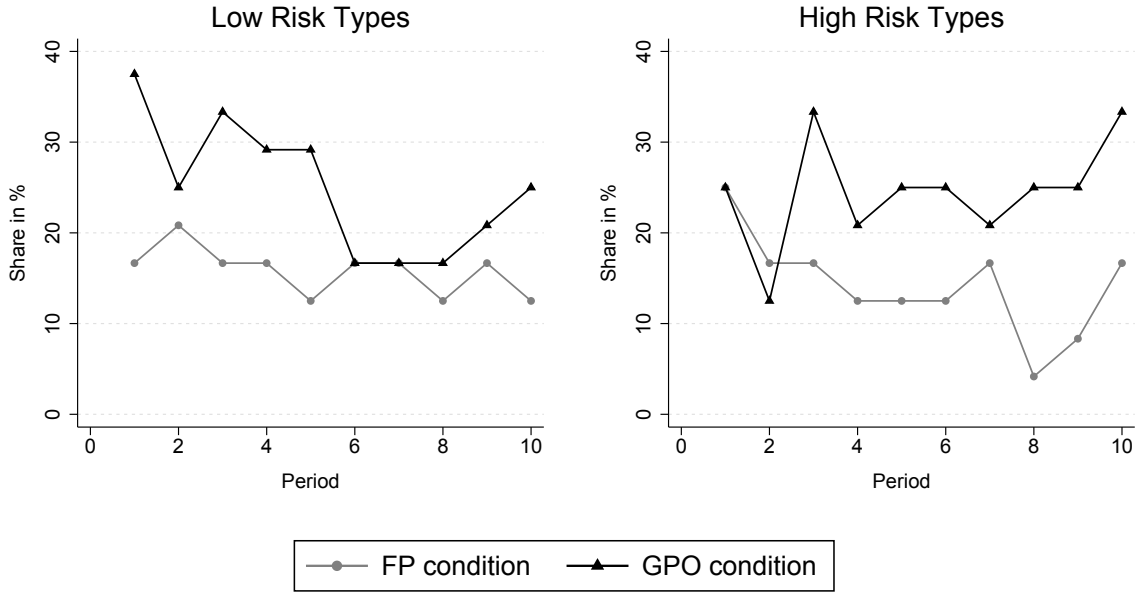

Figure 4.4. Share of Subjects with Positive Net WTP over Time by Genetic Risk Type and Experimental Condition

Notes: The net WTP equals the WTP for group insurance minus the premium for individual insurance.

In what follows, we will focus on subjects with low genetic risk of illness as their willingness to cross-subsidize is crucial for the existence of a mixed group insurance. Table 4.4 shows the estimation results of linear probability models for the propensity of having a positive net WTP.[20] All models include an interaction term between the experimental condition indicator and an indicator for the last five periods to account for the smaller differences in the propensity of having a positive net WTP between the experimental conditions in the last five periods (see Figure 4.4). In the regression table, we report both cluster-robust standard errors and p-values for tests of the null of a zero coefficient computed using the wild cluster bootstrap-t procedure (Cameron et al. 2008). The p-values account for the fact that usual cluster-robust standard errors tend to be downward biased with 12 clusters. Table 4.4 shows, in line with the descriptive results, that for the first five periods the propensity of a positive net WTP is 12.6 to 16.0 percentage points higher in the GPO condition. After controlling for a subject's characteristics[21], the difference between the two experimental conditions is statistically significant at the 5% level. Also in line with the descriptive results, the difference between the two experimental conditions

[20]The following results are robust if we estimate probit instead of linear probability models.

[21]There is an imbalance of economics students across experimental conditions. Despite random assignment of participants to experimental conditions, economics students were 22.5 percentage points more likely to be assigned to the FP condition.

Table 4.4. Regression Results for the Propensity of a Positive Net WTP

	(1)	(2)	(3)
	Dependent variable: Prob(WTP > indiv. premium)		
Constant	0.167**	0.069	-0.064
	(0.067)	(0.698)	(0.619)
	[0.148]	[1.000]	[0.681]
GPO condition	0.142	0.126*	0.160**
	(0.087)	(0.068)	(0.070)
	[0.114]	[0.070]	[0.032]
Last 5 periods	-0.017	-0.017	-0.017
	(0.029)	(0.030)	(0.030)
	[0.642]	[0.642]	[0.642]
GPO condition ×	-0.100*	-0.100*	-0.100*
Last 5 periods	(0.050)	(0.050)	(0.050)
	[0.105]	[0.105]	[0.105]
Male		0.041	0.049
		(0.099)	(0.100)
		[0.721]	[0.665]
Age (in years)		0.004	0.003
		(0.032)	(0.029)
		[0.903]	[0.931]
Economics student		-0.220***	-0.249**
		(0.069)	(0.082)
		[0.028]	[0.036]
Use of health app		0.079	0.084
		(0.114)	(0.104)
		[0.521]	[0.482]
Risk preferences (ref. group: risk averse)			
Risk neutral			0.116
			(0.103)
			[0.360]
Risk seeking			0.102
			(0.071)
			[0.088]
Altruism (ref. group: selfish subjects)			
Low			0.105
			(0.092)
			[0.284]
High			0.176*
			(0.094)
			[0.118]
Adjusted R2	0.02	0.06	0.10
Observations	480	480	480

Notes: Linear probability model estimates. Cluster-robust standard errors are reported in parentheses. Wild cluster-robust p-values, computed using the wild cluster bootstrap-t procedure described in Cameron et al. (2008), are reported in squared brackets. Subjects are classified as risk averse (risk neutral/risk seeking) if they prefer 6 to 10 (4 to 5/0 to 3) times the safe to the risky lottery in the Holt-Laury task. Subjects are classified as selfish (low altruism/high altruism) subject if they choose to donate 0 ECU (1 to 39 ECU/40 to 80 ECU) to charity in the dictator game. The sample is restricted to subjects with low assigned genetic risk of illness.
* p<0.1 ** p<0.05 *** p<0.01.

is smaller in the last five periods. The coefficient on the interaction term is negative, but just lacking statistical significance at the 10% level. The hypothesis that there is no difference across experimental conditions for the last five periods is not rejected at conventional significance levels, however.

The effects of two covariates are worth mentioning. In line with other experimental evidence in the literature (e.g., Bauman and Rose 2011, Frank et al. 1993, Gerlach 2017), Marwell and Ames 1981), students with economics as major make choices that suggest lower social preferences.[22] In our experiment, they are more than 20 percentage points less likely to have a positive net WTP than other students. Moreover, in line with expectations, the propensity of a positive net WTP is positively related to altruism in the incentivized dictator game. Subjects who decided to donate at least half of the 80 ECU to charity were 17.6 percentage points more likely to have a positive net WTP than subjects who decided to donate nothing to charity. This effect is imprecisely estimated, however, and no longer statistically significant after accounting for the small number of clusters.

We observe a large heterogeneity in the willingness to pool across subjects with low genetic risk. Figure 4.5 exemplifies this heterogeneity by displaying WTP over time for four different subjects. Subject 74 is a subject with strong social preferences. Her WTP is substantially larger than her individual insurance premium in all periods. She participates in a mixed group insurance over all ten periods and is willing to cross-subsidize up to three subjects with high genetic risk of illness (periods 5 to 10). Subject 96 displays a moderate willingness to pool. Her WTP is lower than the WTP of subject 74, and in line with conditional participation in group insurance, i.e. participation that depends either on other low genetic risk types' willingness to cross-subsidize or the level of cross-subsidization. In particular, subject 96 is not willing to cross-subsidize more than one high genetic risk type. This is nicely seen from behavior in period 5, when subject 96 returns to her WTP level of period 3 after cross-subsidizing two high genetic risk types in period 4. Subjects 79 and 95 both display no disutility from advantageous inequality resulting from their lower genetic risk. Subject 79 always exactly states her individual insurance premium as WTP, such that she would only be included in a group insurance

[22]When compared to students of other disciplines, economics students made more selfish decisions in a third-party punishment game (Gerlach 2017), were less likely to make donations to social programs (Bauman and Rose 2011), and contributed less of their private savings to the common pot in public goods games (Marwell and Ames 1981). In addition, economists appear to behave less cooperatively in a prisoner's dilemma than non-economists (Frank et al. 1993).

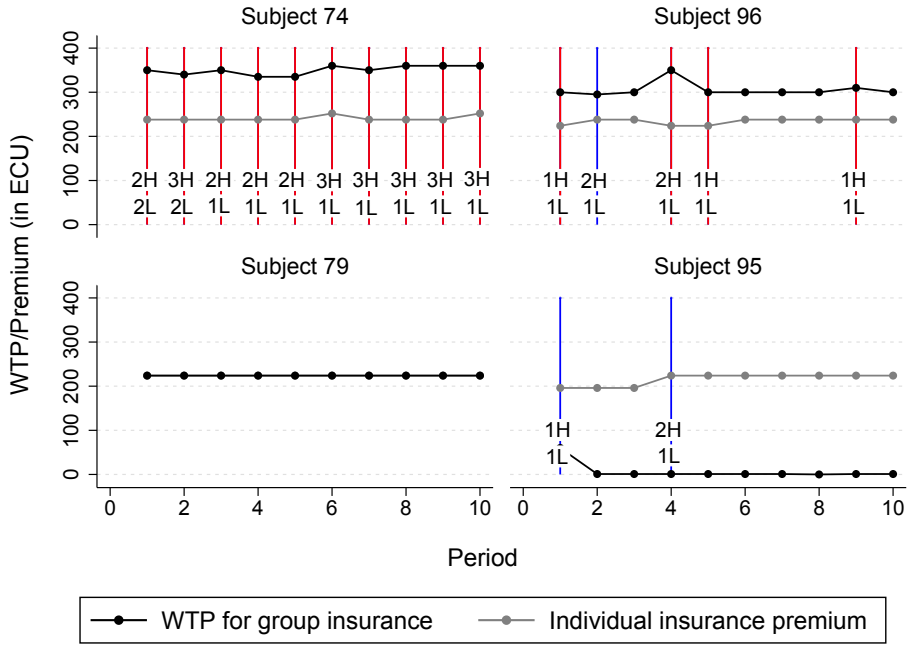

FIGURE 4.5. HETEROGENEITY OF SOCIAL PREFERENCES

Notes: All subjects have a low assigned genetic risk of illness and were assigned to the GPO condition. All subjects prefer individual insurance to no insurance if not included in group insurance. H (L) indicates the number of subjects with high (low) genetic risk who participate in group insurance. Red (blue) vertical lines indicate periods with a mixed group insurance in which the subject (does not) participate(s).

if she was cross-subsidized. Subject 95's choices show that this subject is not willing to pool at all.[23]

How these heterogeneous social preferences translate to existence and level of voluntary pooling of heterogeneous health risks at the societal level depends on the distribution of social preferences of low genetic risk types in a society. For example, a high share of unconditional participants with generally high net WTP (e.g., subject 74) among low genetic risk subjects should lead to more mutualization of genetic risk. Less mutualization should be observed in societies with a high share of conditional participants (e.g., subject 96), and it becomes even more unlikely in these societies if either no unconditional participant is present or conditional participants fail to coordinate. Very little or even no pooling of genetic risk should be observed in societies with a negligible share of low genetic risk subjects who have a positive net WTP in at least some periods. In the next

[23]This conclusion can be drawn even though subject 95 indicates a positive WTP for group insurance in period 1 as her WTP (60 ECU) is far below her smallest possible group insurance premium (266 ECU). Moreover, she reduces her WTP to zero after observing the consequences of being included in a group with one high genetic risk type for another subject with low genetic risk.

section, we will analyze in more depth how the heterogeneity of social preferences at the individual level affects pooling of genetic risk at the societal level.

Before turning to genetic pooling at the societal level, a short look at the role of effort provision and WTP for group insurance across the experimental conditions is interesting. As shown in the overview of results, at the aggregate level, there is no evidence for higher effort in the GPO condition. This goes against our predictions, however, it has to be observed that the level of pooling in both experimental conditions is low such that strong differences cannot arise. At the individual level, we observe that some subjects free-ride on others by providing little effort throughout or reducing their effort upon participation in a group insurance, however. Moreover, we also observe behavior consistent with a reaction to free-riders in the form of a reduction of WTP for group insurance. In the case of subjects with low genetic risk, the reduction in WTP may prevent pooling of genetic risk exposure.

Figure 4.6 exemplifies these reactions for two subjects with low genetic risk in society 2 and two subjects with high genetic risk in society 3. For example, in periods 4 and 5, subject 15 constantly reduces her WTP after being included in a group insurance with low genetic risk subjects who provide less effort than she does. In period 6, she reduces her effort to prevent pooling with free-riders. Only after being no longer included in a group insurance in period 6, subject 15 increases her WTP until she again participates in a mixed group insurance. Subjects 18 and 2 both are free-riders. In general, they provide low effort and indicate a WTP that does not exceed their individual insurance premium. Subject 18, for example, reduces her WTP in period 3 to the level of her individual insurance premium after being included in a non-mixed group insurance and she keeps it at the same level as her individual insurance premium in any period thereafter. By contrast, subject 2 increases her WTP for group insurance and reduces her effort after being included in a non-mixed group insurance for the very first time and paying a lower premium than with individual insurance. Finally, subject 13 represents another subject that reacts to free-riding. In period 3, she strongly reduces her WTP after being pooled with another high genetic risk type and paying a group insurance premium that exceeded her individual insurance premium from period 2. Her WTP for group insurance is close to or even lower than her individual insurance premium in any period after period 2.

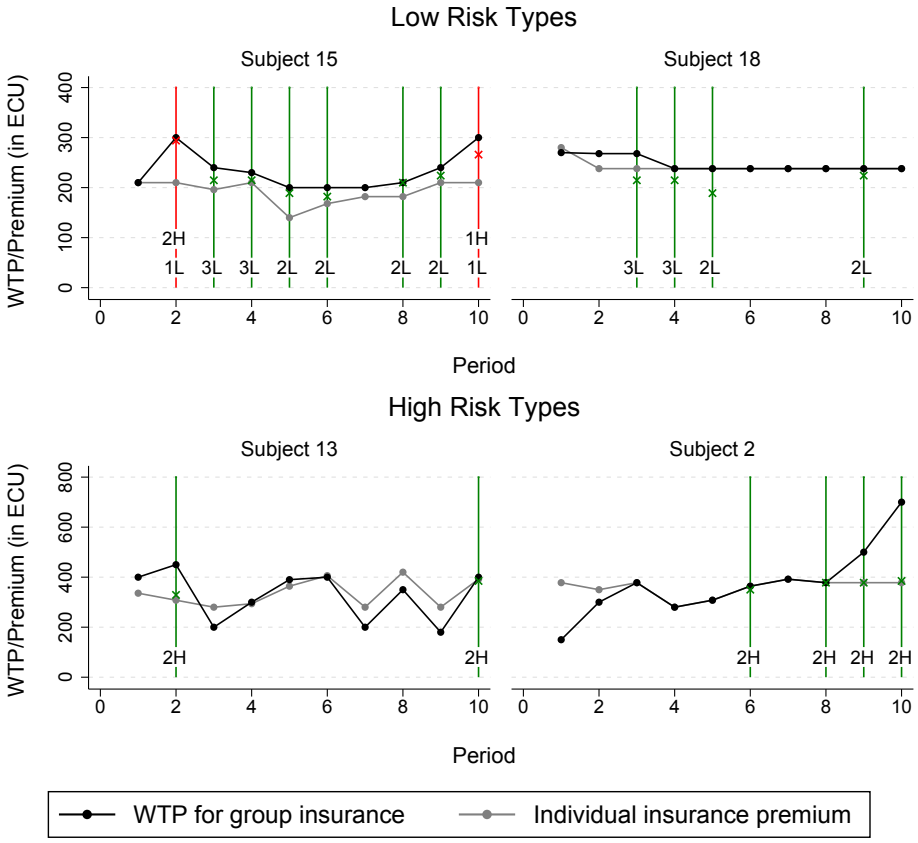

FIGURE 4.6. WTP, EFFORT, AND FREE-RIDING

Notes: All subjects were assigned to the FP condition. Subjects with low (high) genetic risk belong to society 2 (3). H (L) indicates the number of subjects with high (low) genetic risk who participate in group insurance. Red (green) vertical lines indicate periods with a mixed (non-mixed) group insurance in which the subject participates. Red (green) crosses indicate a subject's group insurance premium.

4.3.3 Pooling at the Societal Level

In this section, we analyze in more depth how heterogeneity of social preferences at the individual level affects pooling of genetic risk at the societal level. Figure 4.7 presents the number of subjects with mixed, non-mixed, individual, and no insurance for each of the 12 societies over time. We observe that there exist three types of societies: societies that never experience a mixed group insurance (societies 3, 10, 13, and 20), societies that occasionally experience a mixed group insurance (societies 1, 2, 7, and 16), and societies that often or always experience a mixed group insurance (societies 7, 8, 14, and 17). For societies that never experience a mixed group insurance, the share of low genetic risk subjects who have a positive net WTP in at least one period is at 18.8%, and this share is significantly lower than in societies that experience a mixed group insurance in at least

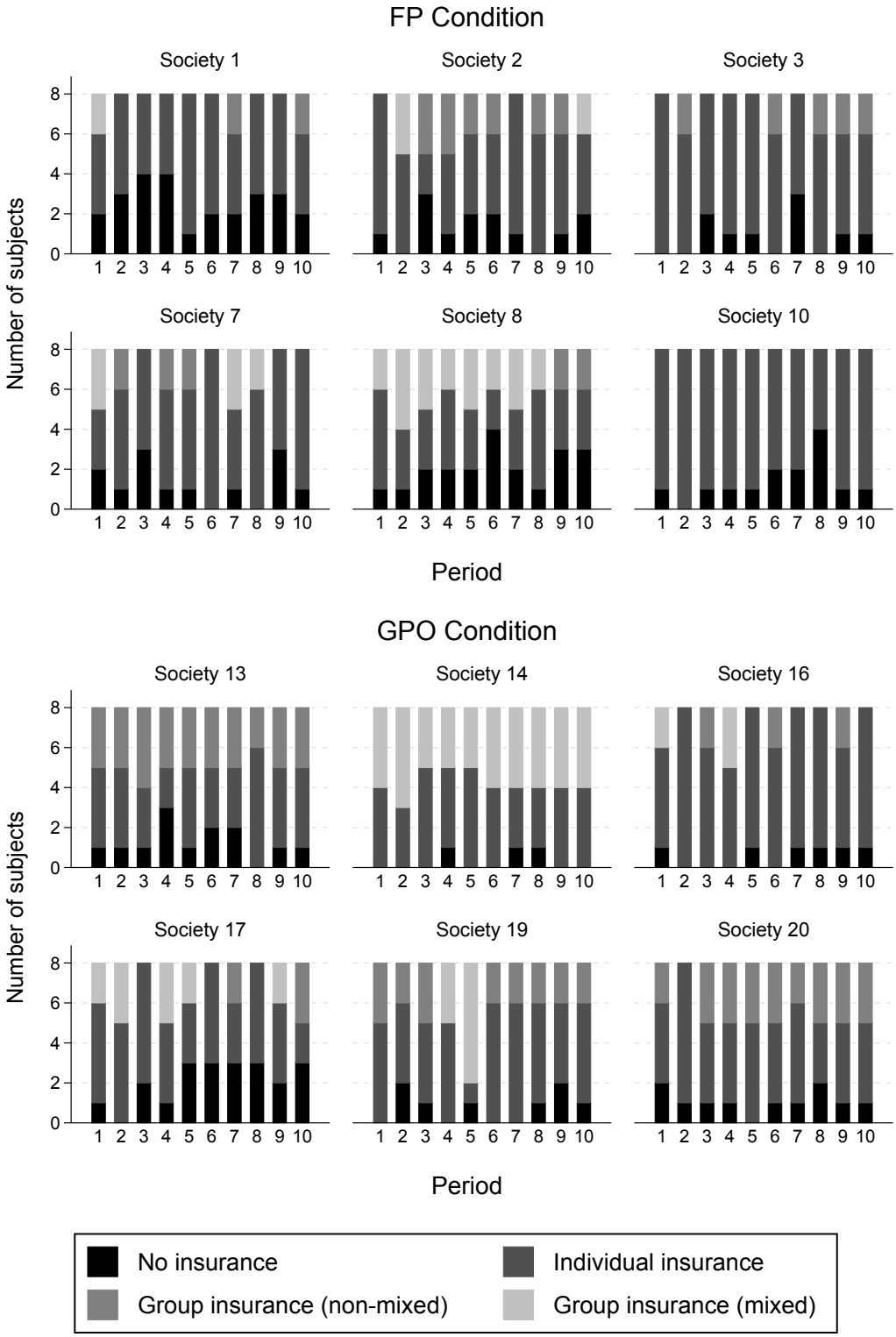

FIGURE 4.7. INSURANCE STATUS OVER TIME BY SOCIETY

one of the ten periods (share for the latter = 65.6%, p-value = 0.0128, Mann-Whitney U-test). Moreover, the share of subjects among low genetic risk subjects who always have a positive net WTP is highest for societies that often or always experience a mixed group insurance. At 18.8%, it is 15.6 percentage points higher than for societies that experience few or no pooling of genetic risk (p-value = 0.0382, Mann-Whitney U-test).

A comparison of societies across experimental conditions reveals that there is a tendency towards more pooling of genetic risk in the GPO condition. Although the share of societies that ever experience a mixed group insurance is balanced across conditions, in the GPO condition societies are more likely to have a mixed group insurance. Moreover, if there is a mixed group insurance, this group insurance has, on average, more members. Both results are, however, not statistically significant. As illustrated by Figure 4A.4 in Appendix 4A, group size in the GPO condition is larger because there more often exist two or three low genetic risk types who are willing to participate in a mixed group insurance in the same period, and their WTP is such that additional high genetic risk types are able to join the group insurance. In addition, group size in the GPO condition is larger because low genetic risk types are more often willing to cross-subsidize more than one high genetic risk type. Overall, these results suggest that, despite strong heterogeneity in social preferences, social preferences are sufficiently strong to allow for some pooling of ex-ante heterogeneous health risks at the societal level. In line with our hypothesis, there is a tendency towards more pooling in the GPO condition.

In order to better understand the group insurance dynamics in a society, we investigated the WTP dynamics of all subjects jointly in a society. In Figure 4.8, we graphically summarize the dynamics for subjects in society 14.[24] For each subject, we show the evolution of her WTP for group insurance and her individual insurance premium, where changes of the latter result from changes of her preventive effort level. Periods with mixed group insurance are indicated by vertical red (if subject participates) and blue (if subject does not participate) lines. In society 14, two subjects with low genetic risk of illness indicate a WTP for group insurance that exceeds their individual insurance premium (subjects 74 and 78). Initially, both subjects contribute to the mixed group insurance, each of them cross-subsidizing one subject with high genetic risk. The increase of sub-

[24]Figure 4A.5 in Appendix 4A provides another example for a society in the FP condition. In this society, there is initially a mixed group insurance, but upon observation of cross-subsidization in period 1 the pivotal low genetic risk type (subject 3) decreases her WTP such that there is no mixed group insurance in any period thereafter.

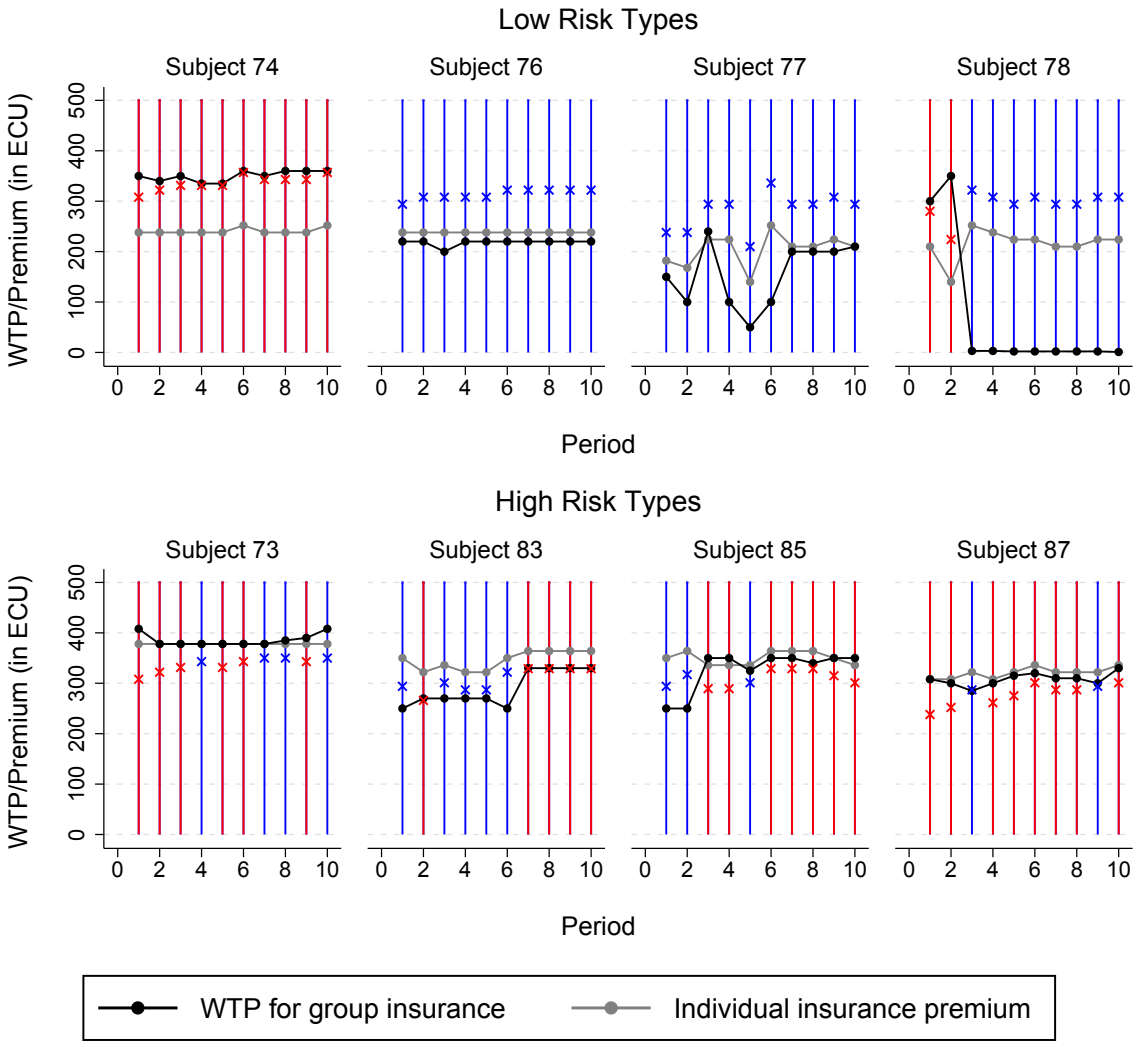

FIGURE 4.8. WTP DYNAMICS OF SOCIETY 14 (GPO CONDITION)

Notes: Red (blue) vertical lines indicate periods with a mixed group insurance in which the subject (does not) participate(s). Red crosses indicate a subject's group insurance premium. Blue crosses indicate a subject's hypothetical group insurance premium, i.e. the group insurance premium that would have had resulted if the subject had participated in the group insurance.

ject 78's WTP for group insurance in period 2 allows one additional subject with high genetic risk to join the group insurance. This higher share of high genetic risk types leads to an increase of the group insurance premium in period 2.[25] Having learned that there is a higher number of high genetic risk types in the group insurance and no third subject with low genetic risk who is willing to participate in a mixed group insurance, subject 78 reduces her WTP to zero from period 3 onwards. Thus, subject 78 appears to be willing to contribute to a group insurance if the share of high genetic risk types is less than two

[25]Note that subject 78's group insurance premium decreases in period 2. This decrease results from a higher effort level of subject 78 in period 2, however. The group insurance premium before effort-related premium discount increases with the addition of one extra high genetic risk type to the group insurance.

third. Given the distribution of social preferences in her society, however, the share of high genetic risk types in group insurance is higher, such that subject 78 reduces her WTP and no longer participates in a group insurance. Mixed group insurance in this society is supported by subject 74, who exhibits strong social preferences, as already discussed in the previous section. Subject 74 decreases her WTP neither upon the participation of one additional high genetic risk type in the group insurance nor upon drop out of subject 78. She eventually increases her WTP in period 6 and keeps it roughly constant after being insured jointly with three high genetic risk subjects.

Overall, we find that sustained mixed group insurance is only possible with the presence of subjects with strong social preferences, and that the dynamics in societies with subjects with moderate WTP for group insurance disfavor the emergence of a mixed group insurance.

4.3.4 Results of the Survey

In order to compare and contrast the results of the incentivized experiment with stated preferences for health insurance systems, we asked subjects which health insurance system they would vote for in a post-experimental survey. In this survey, we also asked subjects whether they were using a health app, and whether they would be willing to share information about their health with their health insurance provider.

Table 4.5 summarizes the results of the post-experimental survey. It shows that more than one quarter of all participants are currently using a health app. With 31.3%, this share is about 50% higher in the FP than in the GPO condition, but this difference is not statistically significant. Moreover, Table 4.5 demonstrates that across experimental conditions 75.0% of the subjects are willing to share health information with their health insurance provider, although the majority of subjects would only do so if they get a premium discount. Differences across experimental conditions are striking and influenced by a subject's condition in the experiment. Subjects in the GPO condition are, for example, 20 percentage points more likely to state that they are willing to share health information with their health insurance provider than subjects in the FP condition (p-value = 0.0184, two-sample test of proportions). Moreover, the share of subjects who would do so if they get a premium discount is at 75.6% ($= \frac{64.4\%}{64.4\%+20.8\%}$) also roughly 20 percentage points

Table 4.5. Results of the Post-Experimental Survey, Shares in %

Variable	FP condition	GPO condition
Use of health app	31.3	22.9
Willingness to share health information		
Yes	29.2	20.8
Yes, if premium discount	35.4	64.6
No	35.4	14.6
Vote on health insurance system		
Individual insurance only	12.5	8.3
Individual + group insurance (GPO)	60.4	62.5
Individual + group insurance (FP)	18.8	1.6
Group insurance (FP) only	8.3	12.5
Observations	48	48

higher in the GPO condition (p-value = 0.0641, two-sample test of proportions), when compared to the FP condition.

Preferences elicited in the post-experimental survey on health insurance systems show that only 12.5% (8.3%) of the subjects in the GPO (FP) condition would vote for a health insurance system with individual insurance only. The majority of subjects (almost two third) would vote for a dual system in which there is available both individual insurance as well as a group insurance that pools genetic health risk but individually prices behavioral health risk. Interestingly, incentivized decisions in the experiment are, however, such that in the societies often only a system with individual insurance only emerges, or a dual system in which different genetic risk types are not pooled in group insurance, such that the genetic part of the premium is identical to the genetic part of the premium in an individual insurance only system.

Focusing on subjects who were assigned a low genetic risk of illness in the experiment, another interesting result emerges if comparing incentivized decisions in the experiment with non-incentivized voting decisions from the survey. Table 4.6 shows that 89.6% ($= \frac{480-10-40}{480}$) of the subjects who were assigned a low genetic risk in the experiment would vote for a health insurance system with group insurance, but that the majority of these subjects ($92.4\% = \frac{237+67+38}{480-10-40} \times 100\%$) is not willing to sacrifice their income in the experiment to support a group insurance that pools health risks, such that the health insurance system with group insurance that they would vote for does not emerge.

Table 4.6. Stated Voting Preferences vs. Incentivized Decisions

	WTP > indiv. premium	
Vote on health insurance system	Yes	No
Individual insurance only	10	40
Individual + group insurance (GPO)	53	237
Individual + group insurance (FP)	23	67
Group insurance (FP) only	12	38

Notes: Subjects with low assigned genetic risk of illness. 480 observations (48 subjects × 10 periods).

4.4 Discussion and Conclusion

Do people voluntarily pool genetic risk in health insurance when they know their own risk profile, and if so, does their willingness to pool genetic risk depend on whether behavioral risk is also pooled or individually priced? Standard economic theory predicts that a dual system with individual insurance and a group insurance that pools genetic risk should not exist when an individual insurance with fully risk-adjusted premium is available because low genetic risk types are never willing to cross-subsidize high genetic risk types. This holds irrespective of whether behavioral risk is separated and priced individually or not. Yet, if people exhibit social preferences in the form of disliking wealth differences that have an origin in uncontrollable heterogeneity in risk exposure, these predictions may no longer hold. Depending on the distribution of social preferences of low genetic risk types in a society, a group insurance or mutual that pools on certain types of risk may well coexist with individually priced health insurance. More pooling of genetic risk should be observed in a mutual system in which group insurance only pools genetic risk and individually prices behavioral risk because free-riding incentives are eliminated in such a system.

In this paper, we used a laboratory experiment to analyze the willingness to pool certain risks, i.e. to cross-subsidize, in health insurance. On a more general level, we tested for social preferences in the health insurance context. As main experimental variation, we varied the degree of risk pooling in a group insurance scheme: In the FP condition, group insurance pooled genetic and behavioral risks whereas, in the GPO condition, group insurance pooled genetic risk but individually priced behavioral risk. Unlike genetic risk of illness, which was either high or low, randomly assigned to subjects, and non-modifiable,

behavioral risk of illness was ex ante identical across subjects but could be reduced by costly preventive effort throughout the experiment.

We find that across experimental conditions half of the subjects who were assigned a low genetic risk of illness exhibit social preferences that manifest in a WTP for group insurance that exceeds their individual insurance premium. Pooled across periods, in the GPO condition low genetic risk subjects were roughly ten percentage points more likely to indicate a WTP that exceeded their individual insurance premium than in the FP condition. This difference across conditions was particularly large in the first five periods. We provide evidence for substantial heterogeneity of social preferences of subjects with low assigned genetic risk of illness and show that the manifestation of social preferences at the societal level, i.e. the existence of a group insurance that pools genetic risk, strongly depends on the low genetic risk subjects' distribution of social preferences in a society. A comparison of societies across experimental condition revealed that more genetic risk pooling occurs in the GPO condition, supporting our main hypothesis. Interestingly, we also find a large discrepancy between stated preferences in a post-experimental survey and revealed preferences in the experiment. Subjects with low assigned genetic risk of illness often stated that they would for a health insurance system with a group insurance that pools genetic risk, but in the experiment less than 8% of these subjects were willing to forgo some of their income in order to participate in a group insurance that pools genetic risk.

The results of this paper are in line with those of Gajdos et al. (2017), who find that 8 to 10% of the subjects with low genetic risk voluntarily participate in a group insurance which pools genetic and behavioral risks. Unlike in this paper, in their study contributions to group insurance are proportional to income, which varies across subjects. Therefore, in their study low genetic risk subjects with low income have an additional incentive to participate in group insurance, which may explain the slightly higher participation shares in their study. This study also relates to the studies of Cappelen et al. (2013) and Mollerstrom et al. (2015), both of which study preferences for redistribution in the context of risky situations. In these studies, the shares of subjects who are willing to redistribute and the actual levels of redistribution are higher than in our study. This may be explained by three crucial differences: First, in their studies, with an explicit ex-post redistribution stage, redistribution is much more salient than in in our study. In our study,

redistribution is implicit in the health insurance scheme, as in many other policy domains. Second, they consider a situation of ex-ante equality in opportunities with redistribution for ex-post income inequalities, i.e. after risk realization, while we consider redistribution that accounts for ex-ante heterogeneity in risk exposure. Third, Mollerstrom et al. (2015) consider choices of spectators rather than stakeholders. Because a spectator's income in the experiment is not affected by her redistributive choice, spectators tend to redistribute more than stakeholders do.

One should be cautious when drawing general conclusions from a single laboratory experiment, but the findings of this paper may have interesting implications for understanding political debates. Our results suggest that mandatory pooling might be needed if, under the veil of ignorance, a society nevertheless wishes to pool certain forms of heterogeneous risk exposure such as genetic risk. The current legislation in most countries implements this implicitly with banning the use of genetic information in health insurance schemes, however, this implicit pooling policy might be under attack for two coalescing reasons: First, good types do have an incentive to advocate voluntary disclosure, and, second, health insurers might find ways to price differentiate implicitly on genetic risk via conditioning on correlated information.

Our experimental set-up used to study the scope of pooling and the role of behavioral risk in health insurance was simple, in particular, the modeling of health risk as additively separable in genetic and behavioral risk. Several interesting extensions could be considered in future work. First, a variation in the extent of cross-subsidization required between different genetic risk types. Second, different marginal costs or benefits of effort for the different genetic risk types, which would decrease or amplify inequality concerns with the separation or inclusion of behavioral risk in the group insurance scheme. Moreover, in our voluntary pooling scheme, the design was close to an opt-in system. It would be interesting to see if more pooling can be achieved if participants have to explicitly opt out of the group insurance scheme.

References

Bauman, Y. and E. Rose (2011): "Selection or Indoctrination: Why Do Economics Students Donate Less than the Rest?" *Journal of Economic Behavior & Organization*, 79, 318–327.

Bolton, G. E. and A. Ockenfels (2000): "ERC: A Theory of Equity, Reciprocity, and Competition," *American Economic Review*, 90, 166–193.

Buckley, N. J., K. Cuff, J. Hurley, L. McLeod, R. Nuscheler, and D. Cameron (2012): "Willingness-to-Pay for Parallel Private Health Insurance: Evidence from a Laboratory Experiment," *Canadian Journal of Economics*, 45, 137–166.

Cameron, A. C., J. B. Gelbach, and D. L. Miller (2008): "Bootstrap-Based Improvements for Inference with Clustered Errors," *Review of Economics and Statistics*, 90, 414–427.

Cappelen, A. W., A. D. Hole, E. Ø. Sørensen, and B. Tungodden (2007): "The Pluralism of Fairness Ideals: An Experimental Approach," *American Economic Review*, 97, 818–827.

Cappelen, A. W., J. Konow, E. Ø. Sørensen, and B. Tungodden (2013): "Just Luck: An Experimental Study of Risk-Taking and Fairness," *American Economic Review*, 103, 1398–1413.

Cappelen, A. W., E. Ø. Sørensen, and B. Tungodden (2010): "Responsibility for What? Fairness and Individual Responsibility," *European Economic Review*, 54, 429–441.

Cherry, T. L., S. Kroll, and J. F. Shogren (2005): "The Impact of Endowment Heterogeneity and Origin on Public Good Contributions: Evidence from the Lab," *Journal of Economic Behavior & Organization*, 57, 357–365.

Engel, C. (2011): "Dictator Games: A Meta Study," *Experimental Economics*, 14, 583–610.

Engelmann, D. and M. Strobel (2004): "Inequality Aversion, Efficiency, and Maximin Preferences in Simple Distribution Experiments," *American Economic Review*, 94, 857–869.

Falk, A., A. Becker, T. Dohmen, D. Huffman, and U. Sunde (2016): "The Preference Survey Module: A Validated Instrument for Measuring Risk, Time, and Social Preferences," IZA Discussion Papers 9674, Institute for the Study of Labor (IZA).

Fehr, E. and K. M. Schmidt (1999): "A Theory of Fairness, Competition, and Cooperation," *Quarterly Journal of Economics*, 114, 817–868.

Fischbacher, U. (2007): "Z-Tree: Zurich Toolbox for Ready-Made Economic Experiments," *Experimental Economics*, 10, 171–178.

Frank, R. H., T. Gilovich, and D. T. Regan (1993): "Does Studying Economics Inhibit Cooperation?" *Journal of Economic Perspectives*, 7, 159–171.

Frohlich, N., J. Oppenheimer, and A. Kurki (2004): "Modeling Other-Regarding Preferences and an Experimental Test," *Public Choice*, 119, 91–117.

Gajdos, T., H. Huber-Yahi, S. Massoni, and J.-C. Vergnaud (2017): "Testing a Boost Policy to Reinforce Mutualized Health Insurance," mimeo.

Gerlach, P. (2017): "The Games Economists Play: Why Economics Students Behave More Selfishly than Other Students," *PLOS ONE*, 12, e0183814.

Holt, C. A. and S. K. Laury (2002): "Risk Aversion and Incentive Effects," *American Economic Review*, 92, 1644–1655.

Kahneman, D., J. L. Knetsch, and R. H. Thaler (1986): "Fairness and the Assumptions of Economics," *Journal of Business*, 59, S285–S300.

Kingsley, D. C. (2016): "Endowment Heterogeneity and Peer Punishment in a Public Good Experiment: Cooperation and Normative Conflict," *Journal of Behavioral and Experimental Economics*, 60, 49–61.

Konow, J. (2000): "Fair Shares: Accountability and Cognitive Dissonance in Allocation Decisions," *American Economic Review*, 90, 1072–1091.

Ku, H. and T. C. Salmon (2013): "Procedural Fairness and the Tolerance for Income Inequality," *European Economic Review*, 64, 111–128.

Marwell, G. and R. E. Ames (1981): "Economists Free Ride, Does Anyone Else?: Experiments on the Provision of Public Goods, IV," *Journal of Public Economics*, 15, 295–310.

Milgrom, P. R. (1981): "Good News and Bad News: Representation Theorems and Applications," *Bell Journal of Economics*, 12, 380–391.

Mollerstrom, J., B.-A. Reme, and E. Ø. Sørensen (2015): "Luck, Choice and Responsibility – An Experimental Study of Fairness Views," *Journal of Public Economics*, 131, 33–40.

Oxoby, R. J. and J. Spraggon (2013): "A Clear and Present Minority: Heterogeneity in the Source of Endowments and the Provision of Public Goods," *Economic Inquiry*, 51, 2071–2082.

Appendix

4A Supplementary Figures

FIGURE 4A.1. PREVENTIVE EFFORT DECISION

Notes: On the left hand side of the screen, subjects observe a picture of the Functional Movement Screen. On the right hand side of the screen, subjects observe their initial endowment and their genetic risk of illness. Moreover, subjects are reminded of the probability of winning a voucher for the Functional Movement Screen (see text below radio buttons), which depends on their preventive effort decision.

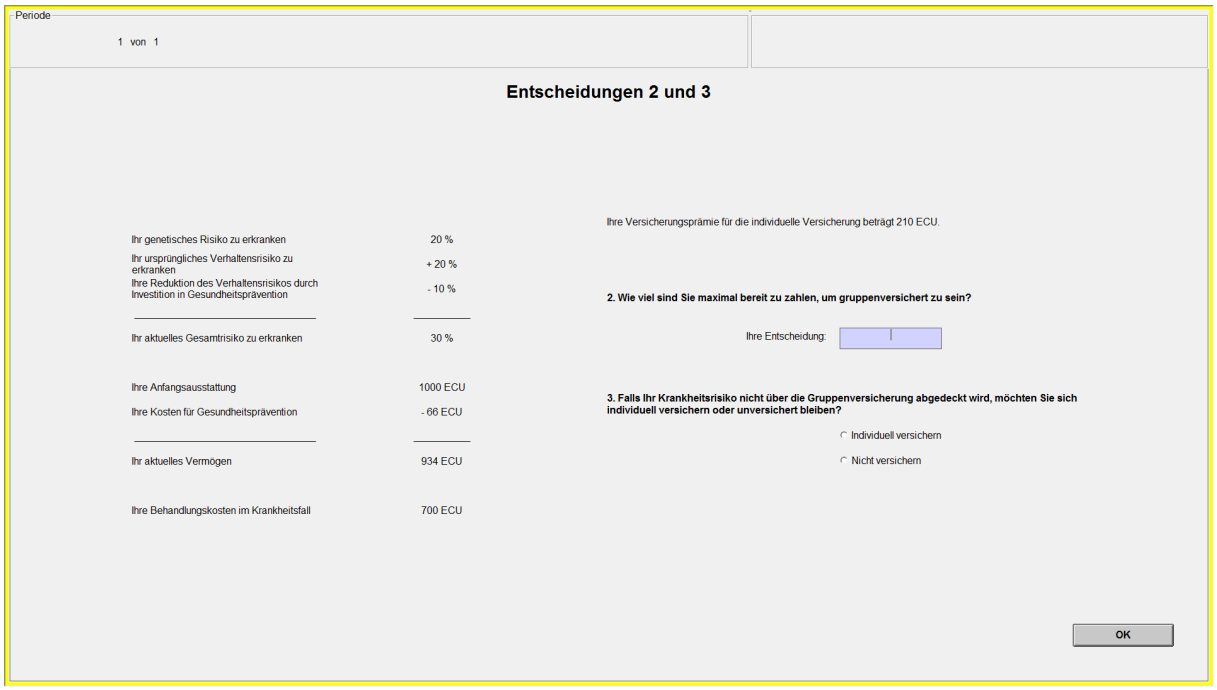

FIGURE 4A.2. HEALTH INSURANCE DECISIONS

Notes: On the left hand side of the screen, subjects observe a summary of the consequences of their preventive effort decision, i.e. subjects observe their overall risk of illness post-prevention and their remaining budget. Moreover, they observe the treatment cost that they would be facing in the case of illness. On the right hand side of the screen, subjects observe their individual insurance premium. Below the insurance premium, they observe the WTP decision for group insurance and the decision between individual and no insurance.

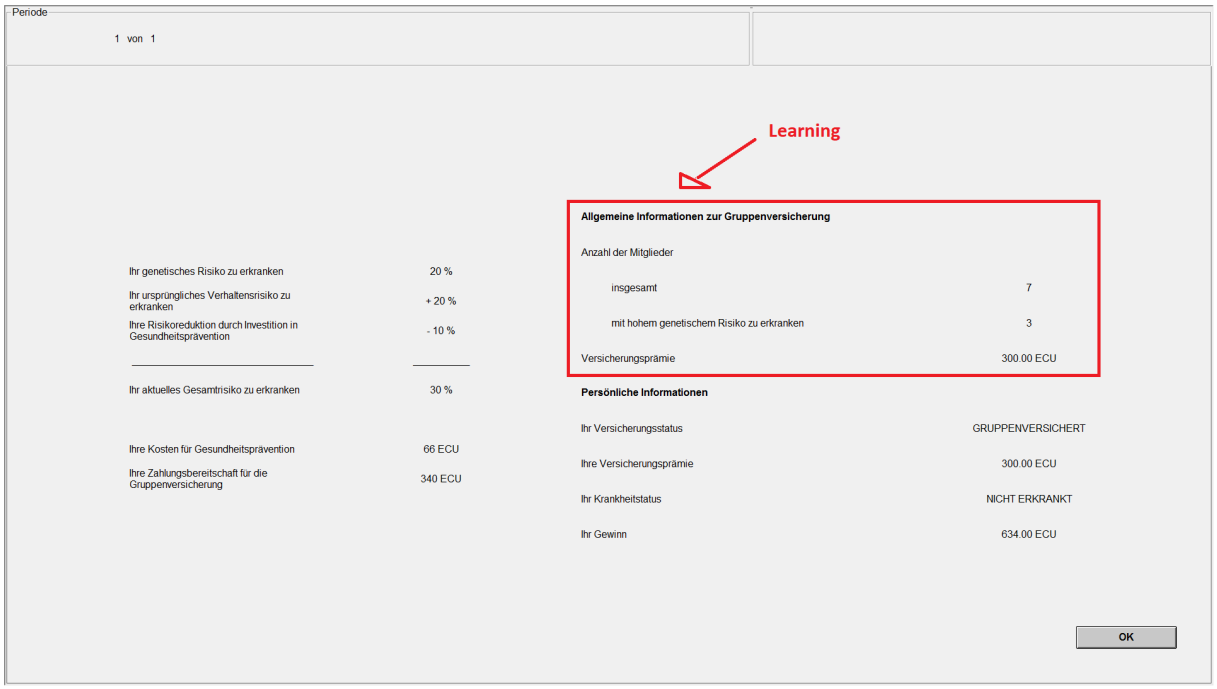

FIGURE 4A.3. SUMMARY SCREEN AT THE END OF A PERIOD

Notes: On the left hand side of the screen, subjects observe a summary of the consequences of their decisions, i.e. subjects observe their overall risk of illness post-prevention, their cost of prevention and their WTP for group insurance. On the right hand side of the screen, subjects observe important end-of-period outcomes. Information on the group insurance comprises the number of group insurance members, the number of group insurance members with high genetic risk, and the group insurance premium. In the GPO condition, the group insurance premium is split up into two parts, a genetic part, which is identical for all group insurance members, and an individual part, which corresponds to the individual premium discount. Private information is given on the insurance status, the paid insurance premium, the illness status, and the final profit.

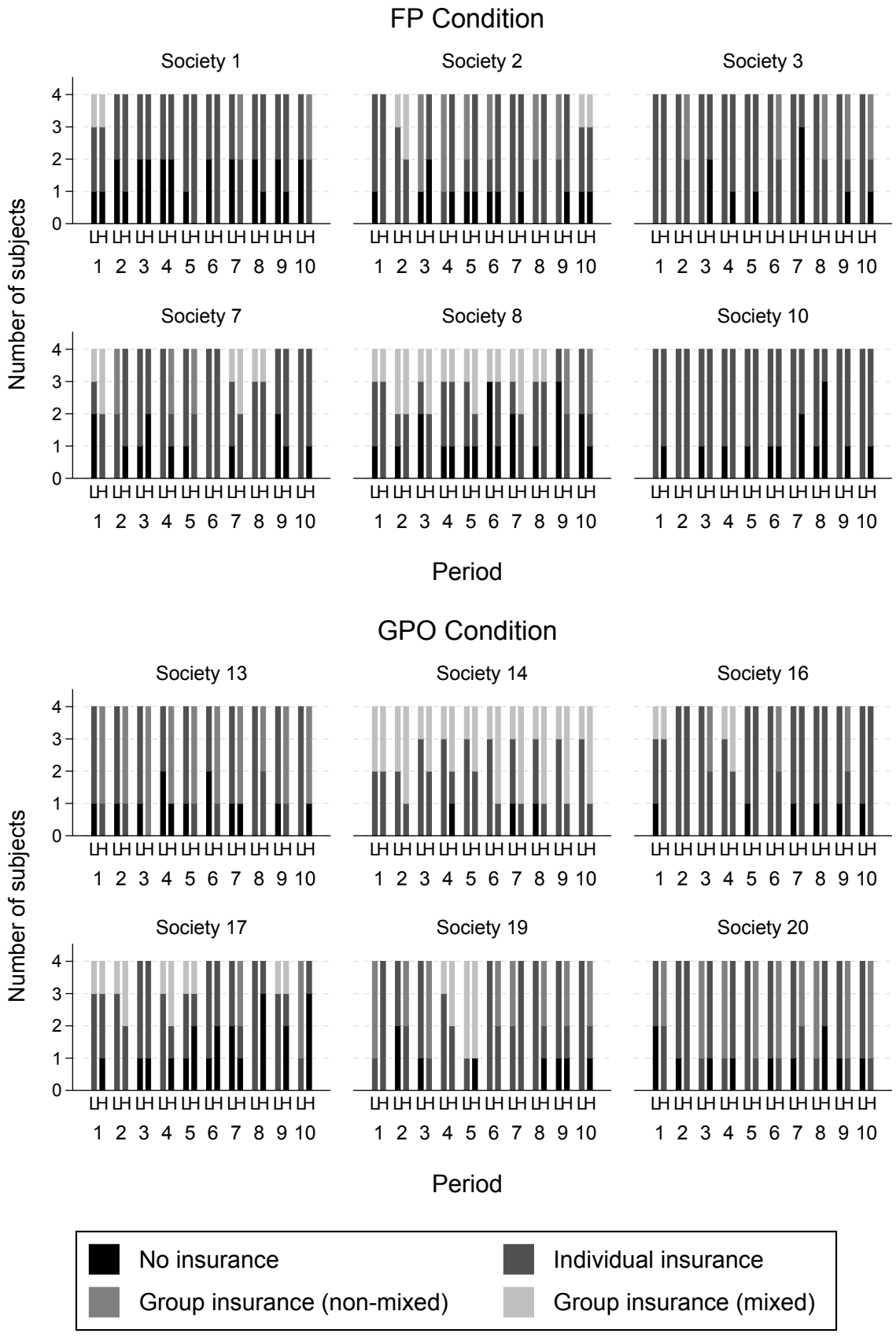

Figure 4A.4. Insurance Status over Time by Genetic Risk Type and Society

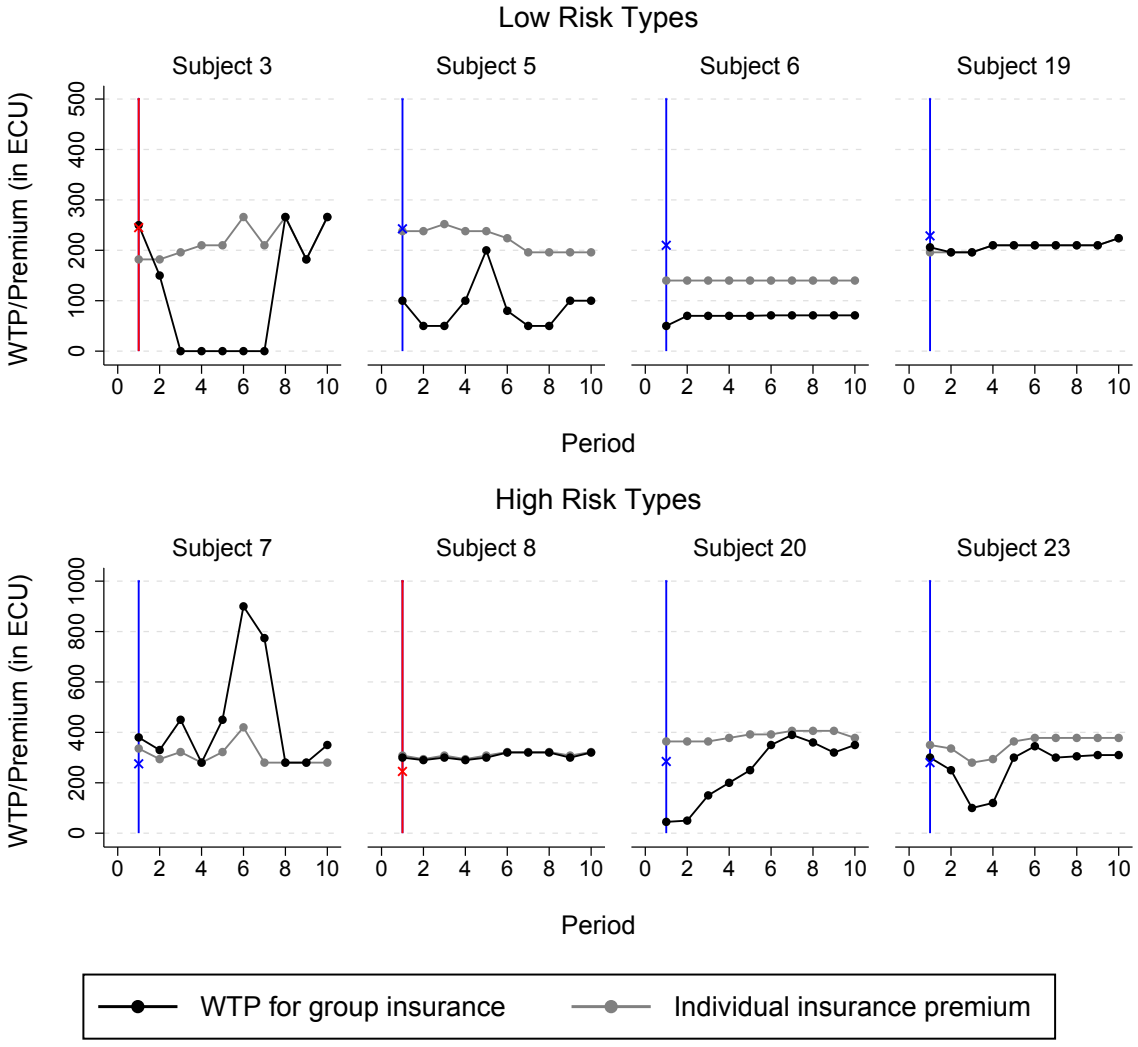

FIGURE 4A.5. WTP DYNAMICS OF SOCIETY 1 (FP CONDITION)

Notes: Red (blue) vertical lines indicate periods with a mixed group insurance in which the subject (does not) participate(s). Red crosses indicate a subject's group insurance premium. Blue crosses indicate a subject's hypothetical group insurance premium, i.e. the group insurance premium that would have had resulted if the subject had participated in the group insurance.

4B Post-Experimental Questionnaire

4B.1 Questionnaire (German Version)

1. Welches Geschlecht haben Sie?
    - Männlich
    - Weiblich
2. Wie alt sind Sie?
3. Sind Sie Student?
    - Ja
    - Nein
4. In welchem Studiengang sind Sie eingeschrieben?
    - Geisteswissenschaften
    - Ingenieurswissenschaften
    - Mathematik
    - Medizin
    - Naturwissenschaften
    - Rechtswissenschaften
    - Theologie
    - Wirtschaftswissenschaften
    - Anderer Studiengang
    - Kein Studiengang
5. Wie schätzen Sie sich selbst ein: Sind Sie eine Person, die prinzipiell bereit ist Risiken einzugehen oder die versucht Risiken zu vermeiden? *Bitte geben Sie auf einer Skala von 0 bis 10 an, wie risikobereit Sie sind. 0 bedeutet, dass Sie keinesfalls Risiken eingehen möchten. 10 bedeutet, dass Sie sehr stark bereit sind Risiken einzugehen.*
6. Wie beurteilen Sie Ihre Bereitschaft mit anderen zu teilen ohne dafür eine Gegenleistung zu erhalten, wenn es sich um eine Wohltätigkeitsorganisation handelt? *Bitte geben Sie auf einer Skala von 0 bis 10 an, wie bereit Sie sind zu teilen. 0 bedeutet, dass Sie keinesfalls bereit sind zu teilen. 10 bedeutet, dass Sie sehr stark bereit sind zu teilen.*
7. Wie schätzen Sie sich selbst ein: Sind Sie eine Person, die grundsätzlich bereit ist unfaires Verhalten zu bestrafen, selbst wenn die Bestrafung mit Kosten für Sie verbunden ist. *Bitte geben Sie auf einer Skala von 0 bis 10 an, wie bereit Sie sind, jemanden trotz Ihrer Kosten zu bestrafen. 0 bedeutet, dass Sie keinesfalls bereit sind trotz Kosten zu bestrafen. 10 bedeutet, dass Sie sehr stark bereit sind trotz Kosten zu bestrafen.*

8. Stellen Sie sich folgende Situation vor: Sie befinden sich in einer unbekannten Stadt und bemerken, dass Sie die Orientierung verloren haben. Sie fragen einen unbekannten Passanten nach dem Weg. Dieser Passant bietet Ihnen an, Sie mit dem Auto zu Ihrem Ziel zu fahren. Die Fahrtzeit beträgt ca. 20 Minuten und kostet den Passanten 20 Franken. Der Passant möchte kein Geld als Gegenleistung erhalten. Sie haben sechs Flaschen Wein bei sich. Die günstigste Flasche kostete 5 Franken, die teuerste 30. Sie entscheiden sich eine der Flaschen an den Passanten als Dankeschön zu geben. Entscheiden Sie sich für die Flasche, die 5, 10, 15, 20, 25, oder 30 CHF gekostet hat?

9. Für welches Gesundheitssystem würden Sie stimmen?

    - Ein Gesundheitssystem, bei dem die Krankenversicherungsprämie des Einzelnen auf dessen individuellem Krankheitsrisiko basiert.
    - Ein Gesundheitssystem, bei dem man wählen kann zwischen einer Gruppenversicherung mit gleicher Krankenversicherungsprämie für alle Gruppenversicherten und einer Versicherung basierend auf dem individuellen Krankheitsrisiko.
    - Ein Gesundheitssystem, bei dem man wählen kann zwischen einer Gruppenversicherung mit einer Krankenversicherungsprämie, die für die Gruppenversicherten genetische Risiken nicht unterscheidet aber Gesundheitsprävention individuell berücksichtigt, und einer Versicherung basierend auf dem individuellen Krankheitsrisiko.
    - Ein Gesundheitssystem mit gleicher Krankenversicherungsprämie für alle.

10. Benutzen Sie eine Gesundheitsapp?

    - Ja
    - Nein

11. Wären Sie bereit Informationen über Ihr Gesundheitsverhalten mit Ihrer Krankenversicherung zu teilen?

    - Ja
    - Ja, aber nur bei Reduktion der Krankenversicherungsprämie
    - Nein

4B.2 Questionnaire (English Translation)

1. What is your gender?
    - Male
    - Female
2. How old are you?
3. Are you a student?
    - Yes
    - No
4. What is your major?
    - Humanities
    - Engineering
    - Mathematics
    - Medicine
    - Natural Science
    - Law
    - Theology
    - Economics
    - Other major
    - No program of study
5. How do you see yourself: Are you a person who is generally willing to take risks or do you try to avoid taking risks? *Please use a scale from 0 to 10, where a 0 means that you are completely unwilling to take risks and a 10 means that you are willing to take risks.*
6. How do you assess your willingness to share with others without expecting anything in return when it comes to charity? *Please use a scale from 0 to 10, where a 0 means that you are completely unwilling to share and a 10 means that you are very willing to share.*
7. How do you see yourself: Are you a person who is generally willing to punish unfair behavior, even if this is costly for you? *Please use a scale from 0 to 10, where a 0 means that you are not willing at all to incur costs to punish unfair behavior and a 10 means you are very willing to incur costs to punish unfair behavior.*
8. Imagine the following situation: You are shopping in an unfamiliar city and you realize that you lost your way. You ask a stranger for directions. The stranger offers to take you with her car to your destination. The ride takes about 20 minutes and costs the stranger about 20 CHF in total. The stranger does not want money for it. You carry six bottles of wine with you. The cheapest bottle costs 5 CHF, the most expensive one 30 CHF. You decide to give one of these bottles to the stranger as a thank-you gift. Which bottle do you give? Do you choose the bottle that costs 5, 10, 15, 20, 25, or 30 CHF?

9. Which health care system would you vote for?
    - A health care system in which the health insurance premium for each individual is based on her own risk of illness.
    - A health care system in which people can choose between a group insurance with identical insurance premium for all group insurance members, and an insurance whose premium is based on the individual risk of illness.
    - A health care system in which people can choose between a group insurance that does not distinguish between genetic risks of illness but individually considers preventive effort, and an insurance whose premium is based on the individual risk of illness.
    - A health care system with identical health insurance premium for everybody.
10. Do you use a health app?
    - Yes
    - No
11. Are you willing to share information about your health with your health insurance provider?
    - Yes
    - Yes, but only if I get a premium discount
    - No

4C Theoretical Predictions: Additional Derivations

4C.1 Utility functions

- Individual insurance:

$$\begin{aligned} U_i^I(e_i) &= y - c(e_i) - P_i^I(e_i) \\ &\quad -\alpha_i \max\left(\Pi_{-i}^{gen,I} - \Pi_i^{gen,I}, 0\right) \\ &\quad -\beta_i \max\left(\Pi_i^{gen,I} - \Pi_{-i}^{gen,I}, 0\right) \\ &= y - c(e_i) - (\pi_i + z - a \cdot e_i) \cdot M \\ &\quad -\alpha_i \max\left((\pi_i - \pi_{-i}) \cdot M, 0\right) \\ &\quad -\beta_i \max\left((\pi_{-i} - \pi_i) \cdot M, 0\right) \qquad \text{for } i = H, L \end{aligned}$$

  and as $\pi_H > \pi_L$:

$$\begin{aligned} U_H^I(e_H) &= y - c(e_H) - (\pi_H + z - a \cdot e_H) \cdot M - \alpha_H \cdot (\pi_H - \pi_L) \cdot M \\ U_L^I(e_L) &= y - c(e_L) - (\pi_L + z - a \cdot e_L) \cdot M - \beta_L \cdot (\pi_H - \pi_L) \cdot M \end{aligned}$$

- Group insurance (GPO condition):

$$\begin{aligned} U_i^G(e_i) &= y - c(e_i) - P_i^G(e_i) \\ &\quad -\alpha_i \max\left(\Pi_{-i}^{gen,G} - \Pi_i^{gen,G}, 0\right) \\ &\quad -\beta_i \max\left(\Pi_i^{gen,G} - \Pi_{-i}^{gen,G}, 0\right) \quad | \quad \Pi_i^{gen,G} = \Pi_{-i}^{gen,G} \\ &= y - c(e_i) - (\bar{\pi} + z - a \cdot e_i) \cdot M \qquad \text{for } i = H, L, \end{aligned}$$

  where $\bar{\pi} = \dfrac{1}{2} \cdot (\pi_H + \pi_L)$

- Group insurance (FP condition):

$$\begin{aligned} U_i^G(e_i, e_{-i}) &= y - c(e_i) - P^G(e_i, e_{-i}) \\ &\quad -\alpha_i \max\left(\Pi_{-i}^{gen,G} - \Pi_i^{gen,G}, 0\right) \\ &\quad -\beta_i \max\left(\Pi_i^{gen,G} - \Pi_{-i}^{gen,G}, 0\right) \quad | \quad \Pi_i^{gen,G} = \Pi_{-i}^{gen,G} \\ &= y - c(e_i) - (\bar{\pi} + z - a \cdot \bar{e}) \cdot M \qquad \text{for } i = H, L, \end{aligned}$$

  where $\bar{\pi} = \dfrac{1}{2} \cdot (\pi_H + \pi_L)$ and $\bar{e} = \dfrac{1}{2} \cdot (e_H + e_L)$

4C.2 Optimal Effort Choices

- Individual insurance:
  - Maximization problem:

$$\max_{e_i} U_i^I(e_i) \quad \text{for } i = H, L$$

  - First order condition:

$$\begin{aligned} \frac{\partial U_i^I(e_i)}{\partial e_i} &= -c'(e_i) + a \cdot M \\ c'(e_i^*) &= a \cdot M \quad \Rightarrow \quad e_H^* = e_L^* = e^* \end{aligned}$$

    An interior solution requires: $c'(0) < a \cdot M < c'(1)$.

  - Second order condition:

$$\frac{\partial^2 U_i^I(e_i)}{\partial e_i^2} = -c''(e_i) < 0 \qquad \text{(due to convexity)}$$

- Group insurance (GPO condition):
  - Maximization problem:

$$\max_{e_i} U_i^G(e_i) \quad \text{for } i = H, L$$

  - First order condition:

$$\begin{aligned} \frac{\partial U_i^G(e_i)}{\partial e_i} &= -c'(e_i) + a \cdot M \\ c'(e_i^*) &= a \cdot M \quad \Rightarrow \quad e_H^* = e_L^* = e^* \end{aligned}$$

    Again, an interior solution requires: $c'(0) < a \cdot M < c'(1)$.

  - Hence, in the GPO condition, optimal effort under group insurance is identical to optimal effort under individual insurance.

- Group insurance (FP condition):
  - Maximization problem:

$$\max_{e_i} U_i^G(e_i, e_{-i}) \quad \text{for } i = H, L$$

– First order condition:

$$\begin{aligned} \frac{\partial U_i^G(e_i, e_{-i})}{\partial e_i} &= -c'(e_i) + \frac{1}{2} \cdot a \cdot M \\ c'(e_i^o) &= \frac{1}{2} \cdot a \cdot M \quad \Rightarrow \quad e_H^o = e_L^o = e^o < e^* \end{aligned}$$

An interior solution requires: $c'(0) < \frac{1}{2} \cdot a \cdot M < c'(1)$.

– Second order condition:

$$\frac{\partial^2 U_i^G(e_i, e_{-i})}{\partial e_i^2} = -c''(e_i) < 0 \qquad \text{(due to convexity)}$$

– In the FP condition, optimal effort under group insurance is smaller than optimal effort under individual insurance. Marginal benefits of effort are reduced by the factor $\frac{1}{2}$ because benefits of effort are split equally among both types.

4C.3 Optimal Insurance Choices: No social preferences ($\alpha_i = \beta_i = 0$)

- GPO condition:

– H-type individual prefers group insurance:

$$\begin{aligned} U_H^G(e^*) &> U_H^I(e^*) \\ y - c(e^*) - (\bar{\pi} - a \cdot e^*) \cdot M &> y - c(e^*) - (\pi_H - a \cdot e^*) \cdot M \\ \frac{1}{2} \cdot (\pi_H + \pi_L) \cdot M &< \pi_H \cdot M \\ \frac{1}{2} \cdot (\pi_L - \pi_H) \cdot M &< 0, \end{aligned}$$

which always holds true for all $M > 0$, since $(\pi_L - \pi_H) < 0$.

– L-type individual prefers individual insurance:

$$\begin{aligned} U_L^I(e^*) &> U_L^G(e^*) \\ y - c(e^*) - (\pi_L - a \cdot e^*) \cdot M &> y - c(e^*) - (\bar{\pi} - a \cdot e^*) \cdot M \\ \pi_L \cdot M &< \frac{1}{2} \cdot (\pi_H + \pi_L) \cdot M \\ 0 &< \frac{1}{2} \cdot (\pi_H - \pi_L) \cdot M, \end{aligned}$$

which always holds true for all $M > 0$, since $(\pi_H - \pi_L) > 0$.

– Since the participation constraint for the L-type individual is violated, group insurance does not exist.

- FP condition:

  - H-type individual will prefer group insurance if

$$\begin{aligned} U_H^I(e^*) &\leq U_H^G(e^o, e^o) \\ \underbrace{\frac{1}{2} \cdot (\pi_L - \pi_H) \cdot M}_{<0} + \underbrace{(e^* - e^o) \cdot a \cdot M}_{>0} &\leq \underbrace{c(e^*) - c(e^o)}_{>0} \end{aligned}$$

  - L-type individual will prefer individual insurance if

$$\begin{aligned} U_L^I(e^*) &> U_L^G(e^o, e^o) \\ \underbrace{\frac{1}{2} \cdot (\pi_H - \pi_L) \cdot M}_{>0} + \underbrace{(e^* - e^o) \cdot a \cdot M}_{>0} &> \underbrace{c(e^*) - c(e^o)}_{>0} \\ a \cdot M &> \frac{c(e^*) - c(e^o)}{e^* - e^o} \end{aligned}$$

    Using the mean value theorem ($c(\cdot)$ cont.), we have

$$a \cdot M > \frac{c(e^*) - c(e^o)}{e^* - e^o} = c'(\xi), \quad \xi \in (e^o, e^*).$$

    This condition always holds for interior solutions because

$$\begin{aligned} e^o < e^* \text{ and } \frac{1}{2} \cdot a \cdot M = c'(e^o) < c'(e^*) = a \cdot M \\ \Rightarrow \frac{1}{2} \cdot a \cdot L < c'(\xi) < a \cdot M, \quad \xi \in (e^o, e^*), \end{aligned}$$

    if $c'(\cdot)$ is continuous and $c''(\cdot) > 0$. It may be violated for corner solutions because then $c'(e^*) > a \cdot M$.

  - Since the participation constraint for the L-type individual is violated, group insurance does not exist.

4C.4 Optimal Insurance Choices: Social Preferences ($\alpha_i \geq \beta_i > 0$, $\beta_i < 1$)

- GPO condition:

  - H-type individual prefers group insurance:

$$U_H^G(e^*) > U_H^I(e^*)$$

$$
\begin{aligned}
y - c(e^*) - (\bar{\pi} + z - a \cdot e^*) \cdot M &> y - c(e^*) - (\pi_H + z - a \cdot e^*) \cdot M \\
&\quad -\alpha_H \cdot (\pi_H - \pi_L) \cdot M \\
\frac{1}{2} \cdot (\pi_H + \pi_L) \cdot M &< \pi_H \cdot M + \alpha_H \cdot (\pi_H - \pi_L) \cdot M \\
-\frac{1}{2} \cdot (\pi_H - \pi_L) \cdot M &< \alpha_H \cdot (\pi_H - \pi_L) \cdot L \\
-\frac{1}{2} &< \alpha_H,
\end{aligned}
$$

which always holds true, since $\alpha_H > 0$.

– L-type individual will prefer individual insurance if

$$
\begin{aligned}
U_L^I(e^*) &> U_L^G(e^*) \\
y - c(e^*) - (\pi_L + z - a \cdot e^*) \cdot M & \\
-\beta_L \cdot (\pi_H - \pi_L) \cdot M &> y - c(e^*) - (\bar{\pi} + z - a \cdot e^*) \cdot M \\
\pi_L \cdot M + \beta_L \cdot (\pi_H - \pi_L) \cdot M &< \frac{1}{2} \cdot (\pi_H + \pi_L) \cdot M \\
\beta_L \cdot (\pi_H - \pi_L) \cdot M &< \frac{1}{2} \cdot (\pi_H - \pi_L) \cdot M \\
\beta_L &< \frac{1}{2}
\end{aligned}
$$

– Hence, there are two cases:

* $\beta_L < \frac{1}{2}$: Group insurance does not exist because the participation constraint for the L-type individual is violated.
* $\beta_L \geq \frac{1}{2}$: Group insurance exists because the L-type individual is sufficiently inequity averse with respect to genetically caused income differences.

- FP condition:

– H-type individual will prefer group insurance if

$$
\begin{aligned}
U_H^G(e^o, e^o) &\geq U_H^I(e^*) \\
y - c(e^o) - (\bar{\pi} + z - a \cdot e^o) \cdot M &\geq y - c(e^*) - (\pi_H + z - a \cdot e^*) \cdot M \\
&\quad -\alpha_H \cdot (\pi_H - \pi_L) \cdot M \\
[a \cdot e^o \cdot M - c(e^o)] - \frac{1}{2} \cdot (\pi_H + \pi_L) \cdot M &\geq [a \cdot e^* \cdot M - c(e^*)] - \pi_H \cdot M \\
&\quad -\alpha_H \cdot (\pi_H - \pi_L) \cdot M
\end{aligned}
$$

$$\begin{aligned}
\frac{1}{2}\cdot(\pi_H-\pi_L)\cdot M+\alpha_H\cdot(\pi_H-\pi_L)\cdot M &\geq [a\cdot e^*\cdot M-c(e^*)]\\
&\quad -[a\cdot e^o\cdot M-c(e^o)]\\
\alpha_H &\geq -\frac{1}{2}+\kappa(e^*,e^o),
\end{aligned}$$

where $\kappa(e^*,e^o)=\dfrac{[a\cdot e^*\cdot M-c(e^*)]-[a\cdot e^o\cdot M-c(e^o)]}{(\pi_H-\pi_L)\cdot M}$.

– L-type individual will prefer individual insurance if

$$\begin{aligned}
U^G(e^o,e^o) &\geq U_L^I(e^*)\\
y-c(e^o)-(\bar{\pi}+z-a\cdot e^o)\cdot M &\geq y-c(e^*)-(\pi_L+z-a\cdot e^*)\cdot M\\
&\quad -\beta_L\cdot(\pi_H-\pi_L)\cdot M\\
[a\cdot e^o\cdot M-c(e^o)]-\frac{1}{2}\cdot(\pi_H+\pi_L)\cdot M &\geq [a\cdot e^*\cdot M-c(e^*)]-\pi_L\cdot M\\
&\quad -\beta_L\cdot(\pi_H-\pi_L)\cdot M\\
-\frac{1}{2}\cdot(\pi_H-\pi_L)\cdot M+\beta_L\cdot(\pi_H-\pi_L)\cdot M &\geq [a\cdot e^*\cdot M-c(e^*)]\\
&\quad -[a\cdot e^o\cdot M-c(e^o)]\\
\beta_L &\geq \frac{1}{2}+\kappa(e^*,e^o).
\end{aligned}$$

– We have that $\kappa(e^*,e^o)>0$ because $f(e_i)=a\cdot e_i\cdot M-c(e_i)$ is maximized at e^* and $(\pi_H-\pi_L)\cdot M>0$. That is, $\kappa(e^*,e^o)$ represents a penalty term that quantifies the efficiency loss resulting from free-riding in the FP condition.

– As $\beta_L<1$ by assumption, the L-type individual is only willing to participate in group insurance, if $\kappa(e^*,e^o)<\frac{1}{2}$. Note that in this case the participation constraint for the H-type individual is always fulfilled.

– Two cases:

  * $\beta_L<\frac{1}{2}+\kappa(e^*,e^o)$: Group insurance does not exist because the participation constraint for group insurance is violated for L-type individual.
  * $\beta_L\geq\frac{1}{2}+\kappa(e^*,e^o)$: Group insurance exists because the L-type individual is sufficiently inequity averse with respect to genetically caused income differences.

4D Instructions

On the following pages, you find the instructions for the experiment:

Anleitung zum Experiment

Herzlichen Dank für Ihre Teilnahme am Experiment. Bitte lesen Sie die folgenden Informationen aufmerksam durch. Falls Sie Fragen zu den Instruktionen haben, heben Sie bitte die Hand. Wir werden dann zu Ihrer Kabine kommen und Ihnen die Fragen beantworten. Bitte sprechen Sie bis zum Ende des Experiments nicht mehr mit anderen Teilnehmern.

Für Ihr rechtzeitiges Erscheinen erhalten Sie 10 Franken. Während des Experiments können Sie weiteres Geld verdienen. Die Höhe Ihres Verdienstes hängt von Ihren Entscheidungen und den Entscheidungen anderer Teilnehmer ab. Sie haben ausserdem die Möglichkeit, einen Gutschein für einen Gesundheitspräventionskurs beim Hochschulsportverein (ASVZ) zu gewinnen. Alle Entscheidungen werden anonym getroffen, d. h. keiner der anderen Teilnehmer erfährt Ihre Identität. Auch die Auszahlung am Ende des Experiments erfolgt anonym, d. h. kein anderer Teilnehmer erhält über Ihre Auszahlung Bescheid. Der Verdienst während des Experiments wird in ECU (=Experimental Currency Unit) angegeben:

$$20 \text{ ECU} = 1 \text{ Franken.}$$

Das Experiment besteht aus **zwei Teilen**:

- Im **ersten** Teil des Experiments treffen Sie über mehrere Perioden hinweg dieselbe Abfolge an Entscheidungen. Zum Ende des Experiments wird **eine dieser Perioden zufällig** ausgewählt und bestimmt Ihren Verdienst aus diesem Teil des Experiments.
- Im **zweiten** Teil des Experiments sehen Sie sich nacheinander verschiedenen Situationen gegenüber, in denen Sie eine oder aber auch mehrere Entscheidungen treffen. Zum Ende des Experiments wird **eine dieser Situationen zufällig** ausgewählt und bestimmt Ihren Verdienst aus diesem Teil des Experiments.

Im Anschluss an das Experiment bitten wir Sie noch einige Fragen zu beantworten. Auf den folgenden Seiten erklären wir den genauen Ablauf des Experiments.

Teil 1

Allgemeine Informationen:

- Teil 1 des Experiments besteht aus 10 Perioden. Innerhalb jeder dieser 10 Perioden treffen Sie dieselbe Abfolge an Entscheidungen.
- In dem Experiment erhalten Sie in jeder Periode eine **Anfangsausstattung** von 1000 ECU und sehen sich dem Risiko ausgesetzt, zu erkranken. Wenn Sie erkranken, führt dies zu **Behandlungskosten** von 700 ECU.
- Ihr Gesamtrisiko zu erkranken setzt sich aus zwei Komponenten zusammen: Einer genetischen Komponente und einer Verhaltenskomponente. *Im Folgenden werden wir einfach von* ***genetischem Risiko*** *und* ***Verhaltensrisiko*** *sprechen.* Ihr genetisches Risiko, d. h. Ihre angeborene Wahrscheinlichkeit zu erkranken, ist entweder niedrig oder hoch. Ihr genetisches Risiko wird Ihnen zufällig zugewiesen und ändert sich während des Experiments nicht. Ihr genetisches Risiko können Sie im Experiment nicht beeinflussen. Ihr ursprüngliches Verhaltensrisiko, d. h. Ihre Wahrscheinlichkeit aufgrund Ihres Verhaltens zu erkranken, beträgt 20%. Ihr Verhaltensrisiko können Sie im Experiment durch Gesundheitsprävention beeinflussen. Ihr Gesamtrisiko zu erkranken ergibt sich aus der Summe von genetischem Risiko und Verhaltensrisiko. Die beiden Komponenten und das resultierende Gesamtrisiko zu erkranken werden Ihnen zu Beginn des Experiments auf Ihrem Bildschirm angezeigt. Diese Informationen sind für andere Experimentteilnehmer nicht sichtbar.

Niedriges Risiko	
Risikokomponente	Wahrscheinlichkeit zu erkranken
Genetisches Risiko	20%
Verhaltensrisiko	20%
Gesamtrisiko	40%

Hohes Risiko	
Risikokomponente	Wahrscheinlichkeit zu erkranken
Genetisches Risiko	40%
Verhaltensrisiko	20%
Gesamtrisiko	60%

- Zu Beginn von Teil 1 werden Sie ausserdem zufällig einer Gesellschaft zugeordnet. Jede Gesellschaft setzt sich aus 8 Teilnehmern zusammen, von denen 4 ein niedriges und 4 ein hohes genetisches Risiko zu erkranken haben. Die Zusammensetzung Ihrer Gesellschaft ändert sich während des Experiments nicht.
- In jeder Periode können Sie grundsätzlich zwei Arten von Entscheidungen treffen: Zum einen eine Gesundheitspräventionsentscheidung, die der Reduktion Ihres Verhaltensrisikos dienen kann, und zum anderen eine Krankenversicherungsentscheidung.

Ablauf einer Periode:

1. Jeder Teilnehmer erhält eine Anfangsausstattung in Höhe von 1000 ECU und sieht sein genetisches Risiko zu erkranken. Jeder Teilnehmer wählt zwischen 11 Leveln, wie viel Gesundheitsprävention er zur Reduktion seines Verhaltensrisikos betreiben möchte. Je höher das Level an Gesundheitsprävention, desto geringer das Verhaltensrisiko zu erkranken und desto höher die Wahrscheinlichkeit, einen Gutschein für ein Präventionsangebot des ASVZs zu erhalten. Bei diesem Präventionsangebot handelt es sich um einen „Functional Movement Screen“ (FMS). Ziel dieses Screens ist es Bewegungsabläufe zu verbessern, um so einer Abnutzung und Schädigung des Bewegungsapparates vorzubeugen. Weitere Informationen zu diesem Screen finden Sie im Abschnitt „Gutscheine für den Functional Movement Screen“.

   Die mit den 11 Leveln verbundenen Reduktionen im Verhaltensrisiko, die Wahrscheinlichkeiten den Gutschein für den FMS zu erhalten und die Kosten für Gesundheitsprävention sehen Sie in folgender Tabelle:

Level an Gesundheitsprävention	Reduktion des Verhaltensrisikos (in Prozentpunkten)	Wahrscheinlichkeit, den Gutschein für den FMS zu erhalten (in Prozent)	Kosten für Gesundheitsprävention (in ECU)
0	0	0	0
1	2	1	8
2	4	2.5	18
3	6	4.5	30
4	8	7	46
5	10	10	66
6	12	13.5	90
7	14	17.5	118
8	16	22	150
9	18	27	186
10	20	33	226

   Das heisst, wenn Sie z. B. ein hohes genetisches Risiko zu erkranken (40%) haben und das Präventionslevel 6 zur Reduktion Ihres Verhaltensrisikos wählen, beträgt Ihr Gesamtrisiko zu erkranken nach Gesundheitsprävention $40\% + 20\% - 12\% = 48\%$. Durch die Gesundheitsprävention erhalten Sie den Gutschein für den FMS beim ASVZ mit einer Wahrscheinlichkeit von 13.5%. Die Kosten für Ihre Gesundheitsprävention belaufen sich auf 90 ECU.

2. Jeder Teilnehmer trifft seine Krankenversicherungsentscheidung. Grundsätzlich gibt es zwei verschiedene Möglichkeiten der Krankenversicherung: entweder eine Gruppenversicherung oder eine individuelle Versicherung. Bei beiden Versicherungen werden im Krankheitsfall die gesamten Behandlungskosten von der Versicherung übernommen. Wenn Sie die Gruppenversicherung haben, dann sind Sie gemeinsam mit anderen Teilnehmern aus Ihrer Gesellschaft versichert. Die Krankenversicherungsprämie wird dann durch Sie **und** die anderen Teilnehmer in der Gruppenversicherung beeinflusst (s. Abschnitt „Prämienberechnung"). Wenn Sie die individuelle Versicherung haben, sind Sie unabhängig von den anderen Gesellschaftsmitgliedern versichert, d. h. in diesem Fall wird die Krankenversicherungsprämie nur durch Sie beeinflusst. Sie können auch unversichert bleiben. Wenn Sie nicht versichert sind, werden Ihnen im Krankheitsfall Ihre Behandlungskosten von der Anfangsausstattung abgezogen.

   Die Krankenversicherungsentscheidung in einer Periode involviert zwei Entscheidungen, die Sie gleichzeitig treffen:

   (i) Jeder Teilnehmer entscheidet, welche Versicherungsprämie er maximal zu zahlen bereit ist, um in der **Gruppenkrankenversicherung** zu sein. Die Gruppenversicherungsprämie ist für alle Gruppenversicherten gleich hoch. Die Zahlungsbereitschaft bestimmt, ob der Teilnehmer über die Gruppenversicherung versichert ist oder nicht. Um am Ende der Periode gruppenversichert zu sein, muss die Zahlungsbereitschaft mindestens der Versicherungsprämie entsprechen, die sich ergibt, wenn der Teilnehmer gruppenversichert wäre. Daher kann es sein, dass ein Teilnehmer nicht gruppenversichert ist, obwohl seine Zahlungsbereitschaft die Gruppenversicherungsprämie übersteigt. Dies wäre z. B. der Fall, wenn die Gruppenversicherungsprämie nach Berücksichtigung des Teilnehmers so stark ansteigen würde, dass diese über der Zahlungsbereitschaft des Teilnehmers liegen würde. Jeder Teilnehmer kann als Zahlungsbereitschaft maximal seine Anfangsausstattung abzüglich seiner Präventionsinvestition (gemäss seines gewählten Levels für Gesundheitsprävention) wählen.

   (ii) Jeder Teilnehmer entscheidet, ob er sein Krankheitsrisiko über die **individuelle Krankenversicherung** abdecken möchte oder **unversichert** bleiben möchte, wenn er aufgrund seiner maximalen Zahlungsbereitschaft für die Gruppenkrankenversicherung nicht gruppenversichert sein sollte.

3. Die Prämie der Gruppenversicherung wird berechnet. Eine Gruppenversicherung wird angeboten, wenn diese im Erwartungswert keinen Verlust macht, d. h. die Summe der Prämien nicht kleiner ist als die erwarteten Behandlungskosten aller Mitglieder einer Gruppenversicherung. Je nach Zahlungsbereitschaften innerhalb einer Gesellschaft kann die Anzahl der Teilnehmer, die gemeinsam in der Gruppenversicherung versichert sind, variieren. So kann es sein, dass z. B. 3 Teilnehmer einer Gesellschaft in der Gruppenversicherung versichert sind und die anderen Teilnehmer der Gesellschaft individuell oder

nicht versichert sind. Wenn keine Gruppenversicherung angeboten werden kann, da im Erwartungswert Verluste entstehen, ist jeder Teilnehmer gemäss seiner Entscheidung entweder individuell versichert oder nicht versichert.

4. Für jeden Teilnehmer entscheidet sich zufällig, ob er in der Periode erkrankt oder nicht. Dabei richtet sich seine Wahrscheinlichkeit zu erkranken nach seinem aktuellen Gesamtrisiko zu erkranken, d. h. seinem Gesamtrisiko nach Reduktion des Verhaltensrisikos.

5. Am Ende der Periode erhält jeder Teilnehmer Informationen zur Gruppenversicherung: Anzahl der Mitglieder insgesamt, Anzahl der Mitglieder mit hohem genetischen Risiko zu erkranken und Versicherungsprämie. Darüber hinaus sieht jeder Teilnehmer seinen Versicherungsstatus (gruppenversichert, individuell versichert, nicht versichert), seine Versicherungsprämie, seinen Krankheitsstatus (erkrankt, nicht erkrankt) und seinen Gewinn.

Ihr Gewinn in einer Periode:

Zu unterscheiden sind vier Fälle:

- Sie sind **gruppenversichert** (Zahlungsbereitschaft $\geq$ Gruppenversicherungsprämie):

  Ihr Gewinn = Ihre Anfangsausstattung - Ihre Kosten für Gesundheitsprävention - Ihre Gruppenversicherungsprämie,

  d. h. Sie zahlen Ihre Kosten für Gesundheitsprävention und Ihre Gruppenversicherungsprämie, währenddessen die Gruppenversicherung im Krankheitsfall Ihre gesamten Behandlungskosten übernimmt.

- Sie sind **individuell versichert** (Zahlungsbereitschaft $<$ Gruppenversicherungsprämie, wenn Sie gruppenversichert wären):

  Ihr Gewinn = Ihre Anfangsausstattung - Ihre Kosten für Gesundheitsprävention - Ihre individuelle Versicherungsprämie,

  d. h. Sie zahlen Ihre Kosten für Gesundheitsprävention und Ihre individuelle Versicherungsprämie, währenddessen die individuelle Versicherung im Krankheitsfall Ihre gesamten Behandlungskosten übernimmt.

- Sie sind **nicht versichert** (Zahlungsbereitschaft < Gruppenversicherungsprämie, wenn Sie gruppenversichert wären):
  - ... und **erkrankt**:

    Ihr Gewinn = Ihre Anfangsausstattung - Ihre Kosten für Gesundheitsprävention - Ihre Behandlungskosten
  - ... und **nicht erkrankt**:

    Ihr Gewinn = Ihre Anfangsausstattung - Ihre Kosten für Gesundheitsprävention,

  d. h. Sie zahlen Ihre Kosten für Gesundheitsprävention und im Krankheitsfall Ihre Behandlungskosten. Eine Versicherungsprämie entfällt.

Prämienberechnung:

- **Individuelle Versicherung:**

  Ihre Prämie = Behandlungskosten × Ihr Gesamtrisiko nach Gesundheitsprävention zu erkranken (in Prozent)/100 ,

  d. h. die individuelle Versicherungsprämie entspricht Ihren erwarteten Behandlungskosten nach Gesundheitsprävention (erwartete Behandlungskosten = Behandlungskosten x aktuelle Wahrscheinlichkeit zu erkranken (in Prozent)/100).

- **Gruppenversicherung:**

  Ihre Prämie = Behandlungskosten × durchschnittliches Gesamtrisiko der Mitglieder der Gruppenversicherung mit Gesundheitsprävention zu erkranken (in Prozent)/100,

  d. h. die Gruppenversicherungsprämie entspricht den erwarteten Behandlungskosten des durchschnittlichen Gruppenversicherten nach Gesundheitsprävention. Im Gegensatz zur individuellen Versicherungsprämie hängt die Gruppenversicherungsprämie nicht nur von Ihrem genetischen Krankheitsrisiko und Ihrer Gesundheitsprävention ab, sondern auch von den genetischen Krankheitsrisiken und der Gesundheitsprävention der anderen Gruppenversicherten.

  Je nach Zahlungsbereitschaften innerhalb einer Gesellschaft kann es mehrere mögliche Gruppenversicherungen in einer Gesellschaft geben, die sich in Bezug auf den Versichertenpool und die Prämie unterscheiden. In diesem Fall wird zunächst die Gruppenversicherung mit der grösstmöglichen Anzahl an Gruppenversicherten gewählt. Sollten dann immer noch mehrere Gruppenversicherungen mit gleicher Versichertenzahl exi-

stieren, so wird innerhalb dieser Gruppenversicherungen die Gruppenversicherung mit der grösstmöglichen Anzahl an Gruppenversicherten mit hohem genetischem Risiko zu erkranken gewählt. Sollten dann immer noch mehrere Gruppenversicherungen mit gleicher Versichertenzahl und gleicher Anzahl an Gruppenversicherten mit hohem genetischem Risiko zu erkranken existieren, so wird eine dieser Gruppenversicherungen zufällig ausgewählt.

Gutscheine für den Functional Movement Screen:

Der „Functional Movement Screen“ ist ein standardisiertes Testverfahren aus Amerika, welches zur Erfassung ineffizienter und schädlicher Bewegungsmuster eingesetzt wird. Hauptziel dieses Screens ist es Bewegungsschwächen zu erkennen und Bewegungsabläufe zu verbessern, um so einer Abnutzung und Schädigung des Bewegungsapparates vorzubeugen. Langfristig sind sowohl die Abnutzung als auch die Schädigung des Bewegungsapparates mit starken Schmerzen verbunden und können zu hohen Behandlungskosten führen (z. B. durch die Behandlung bei einem Orthopäden oder Physiotherapeuten).

Der „Functional Movement Screen“ des ASVZs wird durch ausgebildete Physiotherapeuten angeboten. Er umfasst sieben einfache Bewegungstests zur Quantifizierung von Beweglichkeit, Stabilität und Bewegungsmustern. Die Kosten für diesen Screen belaufen sich auf 60 Franken. Der Zeitaufwand für diesen Screen beträgt 30 Minuten. Termine für diesen Screen sind individuell vereinbar. Weitere Informationen zu diesem Präventionsangebot des ASVZs finden Sie auf der Seite des ASVZs oder erhalten Sie unter der Nummer +41 44 251 60 51.

Wenn Sie den Gutschein erhalten (gemäss der Wahrscheinlichkeit des von Ihnen gewählten Levels der Gesundheitsprävention), deckt der Gutschein die gesamten Kosten dieses Screens.

Teil 2

Informationen zu den verschiedenen Situationen, in denen Sie Entscheidungen treffen, erhalten Sie nach Teil 1 des Experiments auf Ihrem Bildschirm.

Instructions for the Experiment

Thank you for your participation in the experiment. Please read the following information carefully. If you have any questions regarding the instructions, please raise your hand. Then, we will come to your cabin to answer your questions. Please do not talk to other participants any longer until the end of the experiment.

For showing up on time, you will receive 10 Swiss Francs. Throughout the experiment, you can earn more money. The amount of your remuneration depends on your decisions and the decisions of other participants. In addition, you have the opportunity to win a voucher for a health preventative measure, which is offered by the student sports association of the university (ASVZ). All decisions are made anonymously, i.e. none of the other participants learns about your identity. Also the final payoff at the end of the experiment is made anonymously, i.e. none of the other participants is informed about your final payoff. Throughout the experiment, the profits are indicated in ECU (= Experimental Currency Unit):

20 ECU = 1 Swiss Franc.

The experiment consists of **two parts**:

- In the **first** part of the experiment, you will make the same sequence of decisions over several periods. At the end of the experiment, **one of these periods is randomly** selected and determines your payoff in this part of the experiment.
- In the **second** part of the experiment, you will face several distinct situations, in which you have to make one or several decisions. At the end of the experiment, **one of these situations is randomly** selected and determines your payoff in this part of the experiment.

After the experiment, we will still ask you to answer some questions. On the following pages, we will explain the exact procedure of the experiment.

Part 1

General Information:

- Part 1 of the experiment consists of 10 periods. In each of these 10 periods, you make the same sequence of decisions.
- In the experiment, in each period you receive an **initial endowment** of 1000 ECU and you face the risk of illness. If you turn ill, this leads to **treatment cost** of 700 ECU.
- Your overall risk of illness is composed of two components: a genetic component and a behavioral component. *In the following, we will simply refer to these components as* ***genetic risk*** *and* ***behavioral risk****.* Your genetic risk, i.e. your innate probability to turn ill, is either low or high. Your genetic risk is randomly assigned to you and it does not change throughout the experiment. Your genetic risk cannot be influenced by you in the experiment. Your initial behavioral risk, i.e. your probability to turn ill because of your behavior, amounts to 20%. In the experiment, your behavioral risk of illness can be influenced by you through the means of health prevention. Your overall risk of illness is the sum of the genetic and behavioral risk. Both components and the resulting overall risk of illness are shown on your computer screen at the beginning of the experiment. This information is private and not observed by other participants of the experiment.

Low risk	
Risk component	Probability to turn ill
Genetic risk	20%
Behavioral risk	20%
Overall risk	40%

High risk	
Risk component	Probability to turn ill
Genetic risk	40%
Behavioral risk	20%
Overall risk	60%

- At the beginning of part 1, you are also randomly assigned to a society. Each society consists of 8 participants, 4 of whom have a high genetic risk of illness and 4 of whom have a low genetic risk of illness. The composition of your society does not change throughout the experiment.
- In each period, you can make two types of decisions: On the one hand, a health prevention decision, which may serve to reduce your behavioral risk of illness, and on the other hand, a health insurance decision.

Sequence of Events in a Period:

1. Each participant receives an initial endowment of 1000 ECU and observes her genetic risk of illness. Each participant chooses among 11 levels how much health prevention she wants to do in order to reduce her behavioral risk of illness. The higher the level of health prevention, the lower the behavioral risk of illness and the higher the probability to obtain a voucher for a health preventative measure of the ASVZ. This measure comprises a "Functional Movement Screen" (FMS). The goal of this screen is to improve the course of motion in order to prevent degeneration and damage of the musculoskeletal system. You will find more information about this screen in the section "Vouchers for the Functional Movement Screen".

   In the following table, you observe the reduction of behavioral risk, the probability to obtain a voucher for the FMS, and the cost of health prevention for each of the 11 corresponding levels of health prevention:

Level of health prevention	Reduction of behavioral risk (in percentage points)	Probability to obtain the voucher for the FMS (in percent)	Cost of health prevention (in ECU)
0	0	0	0
1	2	1	8
2	4	2.5	18
3	6	4.5	30
4	8	7	46
5	10	10	66
6	12	13.5	90
7	14	17.5	118
8	16	22	150
9	18	27	186
10	20	33	226

   That is, if you, for example, have a high genetic risk of illness (40%) and choose the prevention level 6 to reduce your behavioral risk of illness, then your overall risk of illness after health prevention amounts to $40\% + 20\% - 12\% = 48\%$. Due to the health prevention you obtain the voucher for the FMS at the ASVZ with a probability of 13.5%. The cost of your health prevention amounts to 90 ECU.

2. Each participant makes her health insurance choice. In principle, two possibilities for a health insurance exist: either a group insurance or an individual insurance. Both types

of insurance will pay the full treatment cost in the case of illness. If you have the group insurance, then you will be jointly insured with other participants of your society. The insurance premium will then be influenced by you **and** other participants of the group insurance (cf. section "Premium Calculation"). If you have the individual insurance, you will be insured independently of the other members of your society, i.e. in this case the insurance premium will only be influenced by you. You can also remain uninsured. If you are not insured, the treatment cost will be subtracted from your initial endowment in the case of illness.

The health insurance choice in a given period involves two decisions, which are made simultaneously:

(i) Each participant decides on the insurance premium that she is willing to pay at most in order to join the **group insurance**. The group insurance premium is for all members of the group insurance identical. The willingness to pay determines whether the participant is insured by the group insurance or not. To be group insured at the end of the period, the willingness to pay has to equal at least the insurance premium that would result if the participant was included in the group insurance. Therefore, it may be that a participant is not group insured, even if her willingness to pay exceeds the group insurance premium. This would be the case, for example, if upon consideration of the participant the group insurance premium increased to the extent that it lay above the participant's willingness to pay. Each participant can indicate a willingness to pay which equals at most her initial endowment minus her investment in health prevention (according to her chosen level of health prevention).

(ii) For the case that a participant would not be group insured, given her maximum willingness to pay for group insurance, each participant decides whether she wants to insure her risk of illness by an **individual insurance** or whether she wants to remain **uninsured**.

3. The premium of the group insurance is calculated. A group insurance will be provided if this insurance does not make any losses in expectation, i.e. if the sum of the group insurance premiums is not smaller than the expected treatment costs of all group insurance members. Depending on the willingness to pay in a society, the number of participants who are jointly insured by the group insurance may vary. It may be, for example, that 3 participants of a society are group insured while the other participants of the society are individually insured or not insured. If no group insurance is provided as the group insurance would make losses in expectation, each participant is either individually insured or not insured according to her decision.

4. For each participant, it is randomly determined whether she turns ill in the period or not. The probability to turn ill corresponds to her actual overall risk of illness, i.e. her overall risk after the reduction of her behavioral risk.

5. At the end of the period, each participant obtains some information about the group insurance: number of members in total, number of members with high genetic risk of illness, and insurance premium. Moreover, each participant observes her insurance status (group insured, individually insured, not insured), her insurance premium, her illness status (ill, not ill), and her profit.

Your Profit in a Period:

There are four cases, which need to be distinguished:

- You are **group insured** (willingness to pay $\geq$ group insurance premium):

  Your profit = your initial endowment - your cost of health prevention - your group insurance premium,

  i.e. you pay your cost of health prevention and your group insurance premium, while the group insurance covers your full treatment cost in the case of illness.

- You are **individually insured** (willingness to pay $<$ group insurance premium if you were included in the group insurance):

  Your profit = your initial endowment - your cost of health prevention - your individual insurance premium,

  i.e. you pay the cost of health prevention and your individual insurance premium, while the individual insurance covers your full treatment cost in the case of illness.

- You are **not insured** (willingness to pay $<$ group insurance premium if you were included in the group insurance):

  - ... and **ill**:

    Your profit = your initial endowment - your cost of health prevention - your treatment cost

  - ... and **not ill**:

    Your profit = your initial endowment - your cost of health prevention,

  i.e. you pay your cost of health prevention and, in the case of illness, your treatment cost. An insurance premium does not need to be paid.

Premium Calculation:

- **Individual insurance:**

  Your premium = treatment cost × your overall risk of illness after health prevention (in percent)/100,

  i.e. the individual insurance premium corresponds to your expected treatment cost after health prevention (expected treatment cost = treatment cost × actual probability to turn ill (in percent)/100).

- **Group insurance:**

  Your premium = treatment cost × average overall risk of illness of all group insurance members after health prevention (in percent)/100,

  i.e. the group insurance premium corresponds to the expected treatment cost of the average group insurance member after health prevention. In contrast to the individual insurance premium, the group insurance premium depends not only on your genetic risk of illness and your health prevention, but also on the genetic risks of illness and the health prevention of the other group insurance members.

  Depending on the willingness to pay in a society, there may exist several possible group insurances in a society, which differ with respect to the pool of insured and the premium. In this case, the group insurance with the highest number of group insurance members is selected first. If there still exist several group insurances with the same number of group insurance members, the group insurance with the highest number of group insurance members with high genetic risk is selected among the remaining ones next. If there still exist several group insurances with the same number of group insurance members and the same number of group insurance members with high genetic risk of illness, one of the remaining group insurances is selected at random.

Vouchers for the Functional Movement Screen:

The "Functional Movement Screen" is a standardized test procedure that was developed in the United States and is used to detect inefficient and harmful movement patterns. The primary goal of this screen is that of detecting weaknesses in movement orders and impoving the course of motion in order to prevent degeneration and damage of the musculoskeletal system. In the long run, degeneration as well as damage of the musculoskeletal system causes strong pain and may lead to high treatment costs (e.g., due to the treatment by an orthopedic specialist or a physiotherapist).

The "Functional Movement Screen" at the ASVZ is offered by professionally trained physiotherapists. It comprises seven simple movement tests to quantify mobility, stability, and movement patterns. The cost of this screening amounts to 60 Swiss Francs. The time required for this screening is 30 minutes. Appointments for this screening are made individually. More information about this health preventative measure of the ASVZ is found on the webpage of the ASVZ or is obtained by dialing the number +41 44 251 60 51.

If you receive the voucher (according to the probability that is attached to your chosen level of health prevention), this voucher will cover the full cost of this screening.

Part 2

After completion of part 1 of the experiment, you will receive further information about the distinct decision situations, which you will be facing, on your computer screen.

Anleitung zum Experiment

Herzlichen Dank für Ihre Teilnahme am Experiment. Bitte lesen Sie die folgenden Informationen aufmerksam durch. Falls Sie Fragen zu den Instruktionen haben, heben Sie bitte die Hand. Wir werden dann zu Ihrer Kabine kommen und Ihnen die Fragen beantworten. Bitte sprechen Sie bis zum Ende des Experiments nicht mehr mit anderen Teilnehmern.

Für Ihr rechtzeitiges Erscheinen erhalten Sie 10 Franken. Während des Experiments können Sie weiteres Geld verdienen. Die Höhe Ihres Verdienstes hängt von Ihren Entscheidungen und den Entscheidungen anderer Teilnehmer ab. Sie haben ausserdem die Möglichkeit, einen Gutschein für einen Gesundheitspräventionskurs beim Hochschulsportverein (ASVZ) zu gewinnen. Alle Entscheidungen werden anonym getroffen, d. h. keiner der anderen Teilnehmer erfährt Ihre Identität. Auch die Auszahlung am Ende des Experiments erfolgt anonym, d. h. kein anderer Teilnehmer erhält über Ihre Auszahlung Bescheid. Der Verdienst während des Experiments wird in ECU (=Experimental Currency Unit) angegeben:

$$20 \text{ ECU} = 1 \text{ Franken.}$$

Das Experiment besteht aus **zwei Teilen**:

- Im **ersten** Teil des Experiments treffen Sie über mehrere Perioden hinweg dieselbe Abfolge an Entscheidungen. Zum Ende des Experiments wird **eine dieser Perioden zufällig** ausgewählt und bestimmt Ihren Verdienst aus diesem Teil des Experiments.
- Im **zweiten** Teil des Experiments sehen Sie sich nacheinander verschiedenen Situationen gegenüber, in denen Sie eine oder aber auch mehrere Entscheidungen treffen. Zum Ende des Experiments wird **eine dieser Situationen zufällig** ausgewählt und bestimmt Ihren Verdienst aus diesem Teil des Experiments.

Im Anschluss an das Experiment bitten wir Sie noch einige Fragen zu beantworten. Auf den folgenden Seiten erklären wir den genauen Ablauf des Experiments.

Teil 1

Allgemeine Informationen:

- Teil 1 des Experiments besteht aus 10 Perioden. Innerhalb jeder dieser 10 Perioden treffen Sie dieselbe Abfolge an Entscheidungen.
- In dem Experiment erhalten Sie in jeder Periode eine **Anfangsausstattung** von 1000 ECU und sehen sich dem Risiko ausgesetzt, zu erkranken. Wenn Sie erkranken, führt dies zu **Behandlungskosten** von 700 ECU.
- Ihr Gesamtrisiko zu erkranken setzt sich aus zwei Komponenten zusammen: Einer genetischen Komponente und einer Verhaltenskomponente. *Im Folgenden werden wir einfach von* ***genetischem Risiko*** *und* ***Verhaltensrisiko*** *sprechen.* Ihr genetisches Risiko, d. h. Ihre angeborene Wahrscheinlichkeit zu erkranken, ist entweder niedrig oder hoch. Ihr genetisches Risiko wird Ihnen zufällig zugewiesen und ändert sich während des Experiments nicht. Ihr genetisches Risiko können Sie im Experiment nicht beeinflussen. Ihr ursprüngliches Verhaltensrisiko, d. h. Ihre Wahrscheinlichkeit aufgrund Ihres Verhaltens zu erkranken, beträgt 20%. Ihr Verhaltensrisiko können Sie im Experiment durch Gesundheitsprävention beeinflussen. Ihr Gesamtrisiko zu erkranken ergibt sich aus der Summe von genetischem Risiko und Verhaltensrisiko. Die beiden Komponenten und das resultierende Gesamtrisiko zu erkranken werden Ihnen zu Beginn des Experiments auf Ihrem Bildschirm angezeigt. Diese Informationen sind für andere Experimentteilnehmer nicht sichtbar.

Niedriges Risiko	
Risikokomponente	Wahrscheinlichkeit zu erkranken
Genetisches Risiko	20%
Verhaltensrisiko	20%
Gesamtrisiko	40%

Hohes Risiko	
Risikokomponente	Wahrscheinlichkeit zu erkranken
Genetisches Risiko	40%
Verhaltensrisiko	20%
Gesamtrisiko	60%

- Zu Beginn von Teil 1 werden Sie ausserdem zufällig einer Gesellschaft zugeordnet. Jede Gesellschaft setzt sich aus 8 Teilnehmern zusammen, von denen 4 ein niedriges und 4 ein hohes genetisches Risiko zu erkranken haben. Die Zusammensetzung Ihrer Gesellschaft ändert sich während des Experiments nicht.
- In jeder Periode können Sie grundsätzlich zwei Arten von Entscheidungen treffen: Zum einen eine Gesundheitspräventionsentscheidung, die der Reduktion Ihres Verhaltensrisikos dienen kann, und zum anderen eine Krankenversicherungsentscheidung.

Ablauf einer Periode:

1. Jeder Teilnehmer erhält eine Anfangsausstattung in Höhe von 1000 ECU und sieht sein genetisches Risiko zu erkranken. Jeder Teilnehmer wählt zwischen 11 Leveln, wie viel Gesundheitsprävention er zur Reduktion seines Verhaltensrisikos betreiben möchte. Je höher das Level an Gesundheitsprävention, desto geringer das Verhaltensrisiko zu erkranken und desto höher die Wahrscheinlichkeit, einen Gutschein für ein Präventionsangebot des ASVZs zu erhalten. Bei diesem Präventionsangebot handelt es sich um einen „Functional Movement Screen“ (FMS). Ziel dieses Screens ist es Bewegungsabläufe zu verbessern, um so einer Abnutzung und Schädigung des Bewegungsapparates vorzubeugen. Weitere Informationen zu diesem Screen finden Sie im Abschnitt „Gutscheine für den Functional Movement Screen“.

   Die mit den 11 Leveln verbundenen Reduktionen im Verhaltensrisiko, die Wahrscheinlichkeiten den Gutschein für den FMS zu erhalten und die Kosten für Gesundheitsprävention sehen Sie in folgender Tabelle:

Level an Gesundheitsprävention	Reduktion des Verhaltensrisikos (in Prozentpunkten)	Wahrscheinlichkeit, den Gutschein für den FMS zu erhalten (in Prozent)	Kosten für Gesundheitsprävention (in ECU)
0	0	0	0
1	2	1	8
2	4	2.5	18
3	6	4.5	30
4	8	7	46
5	10	10	66
6	12	13.5	90
7	14	17.5	118
8	16	22	150
9	18	27	186
10	20	33	226

   Das heisst, wenn Sie z. B. ein hohes genetisches Risiko zu erkranken (40%) haben und das Präventionslevel 6 zur Reduktion Ihres Verhaltensrisikos wählen, beträgt Ihr Gesamtrisiko zu erkranken nach Gesundheitsprävention $40\% + 20\% - 12\% = 48\%$. Durch die Gesundheitsprävention erhalten Sie den Gutschein für den FMS beim ASVZ mit einer Wahrscheinlichkeit von 13.5%. Die Kosten für Ihre Gesundheitsprävention belaufen sich auf 90 ECU.

2. Jeder Teilnehmer trifft seine Krankenversicherungsentscheidung. Grundsätzlich gibt es zwei verschiedene Möglichkeiten der Krankenversicherung: entweder eine Gruppenversicherung oder eine individuelle Versicherung. Bei beiden Versicherungen werden im Krankheitsfall die gesamten Behandlungskosten von der Versicherung übernommen. Wenn Sie die Gruppenversicherung haben, dann sind Sie gemeinsam mit anderen Teilnehmern aus Ihrer Gesellschaft versichert. Die Krankenversicherungsprämie wird dann durch Sie **und** die anderen Teilnehmer in der Gruppenversicherung beeinflusst (s. Abschnitt „Prämienberechnung"). Wenn Sie die individuelle Versicherung haben, sind Sie unabhängig von den anderen Gesellschaftsmitgliedern versichert, d. h. in diesem Fall wird die Krankenversicherungsprämie nur durch Sie beeinflusst. Sie können auch unversichert bleiben. Wenn Sie nicht versichert sind, werden Ihnen im Krankheitsfall Ihre Behandlungskosten von der Anfangsausstattung abgezogen.

   Die Krankenversicherungsentscheidung in einer Periode involviert zwei Entscheidungen, die Sie gleichzeitig treffen:

   (i) Jeder Teilnehmer entscheidet, welche Versicherungsprämie er maximal zu zahlen bereit ist, um in der **Gruppenkrankenversicherung** zu sein. Die Gruppenversicherungsprämie setzt sich aus zwei Teilen zusammen: einem Teil, der für alle Gruppenversicherten gleich ist, und einem individuellen Teil, der die Prämienreduktion durch Gesundheitsprävention berücksichtigt:

   – Den erwarteten Behandlungskosten des durchschnittlichen Gruppenversicherten vor Gesundheitsprävention
   – **abzüglich** Ihrer Prämienreduktion durch Gesundheitsprävention.

   Die Zahlungsbereitschaft bestimmt, ob der Teilnehmer über die Gruppenversicherung versichert ist oder nicht. Um am Ende der Periode gruppenversichert zu sein, muss die Zahlungsbereitschaft mindestens der Versicherungsprämie entsprechen, die sich insgesamt ergibt, wenn der Teilnehmer gruppenversichert wäre. Daher kann es sein, dass ein Teilnehmer nicht gruppenversichert ist, obwohl seine Zahlungsbereitschaft die gesamte Versicherungsprämie in der Gruppenversicherung übersteigt. Dies wäre z. B. der Fall, wenn die gesamte Gruppenversicherungsprämie nach Berücksichtigung des Teilnehmers so stark ansteigen würde, dass diese über der Zahlungsbereitschaft des Teilnehmers liegen würde. Jeder Teilnehmer kann als Zahlungsbereitschaft maximal seine Anfangsausstattung abzüglich seiner Präventionsinvestition (gemäss seines gewählten Levels für Gesundheitsprävention) wählen.

   (ii) Jeder Teilnehmer entscheidet, ob er sein Krankheitsrisiko über die **individuelle Krankenversicherung** abdecken möchte oder **unversichert** bleiben möchte, wenn er aufgrund seiner maximalen Zahlungsbereitschaft für die Gruppenkrankenversicherung nicht gruppenversichert sein sollte.

3. Die Prämien der Gruppenversicherung werden berechnet. Eine Gruppenversicherung wird angeboten, wenn diese im Erwartungswert keinen Verlust macht, d. h. die Summe der Prämien nicht kleiner ist als die erwarteten Behandlungskosten aller Mitglieder einer Gruppenversicherung. Je nach Zahlungsbereitschaften innerhalb einer Gesellschaft kann die Anzahl der Teilnehmer, die gemeinsam in der Gruppenversicherung versichert sind, variieren. So kann es sein, dass z. B. 3 Teilnehmer einer Gesellschaft in der Gruppenversicherung versichert sind und die anderen Teilnehmer der Gesellschaft individuell oder nicht versichert sind. Wenn keine Gruppenversicherung angeboten werden kann, da im Erwartungswert Verluste entstehen, ist jeder Teilnehmer gemäss seiner Entscheidung entweder individuell versichert oder nicht versichert.

4. Für jeden Teilnehmer entscheidet sich zufällig, ob er in der Periode erkrankt oder nicht. Dabei richtet sich seine Wahrscheinlichkeit zu erkranken nach seinem aktuellen Gesamtrisiko zu erkranken, d. h. seinem Gesamtrisiko nach Reduktion des Verhaltensrisikos.

5. Am Ende der Periode erhält jeder Teilnehmer Informationen zur Gruppenversicherung: Anzahl der Mitglieder insgesamt, Anzahl der Mitglieder mit hohem genetischen Risiko zu erkranken, Versicherungsprämie (vor Berücksichtigung der Prämienreduktion durch Gesundheitsprävention) und seine Prämienreduktion durch Gesundheitsprävention. Darüber hinaus sieht jeder Teilnehmer seinen Versicherungsstatus (gruppenversichert, individuell versichert, nicht versichert), seine Versicherungsprämie (nach Berücksichtigung der Prämienreduktion durch Gesundheitsprävention), seinen Krankheitsstatus (erkrankt, nicht erkrankt) und seinen Gewinn.

Ihr Gewinn in einer Periode:

Zu unterscheiden sind vier Fälle:

- Sie sind **gruppenversichert** (Zahlungsbereitschaft $\geq$ Ihre Gruppenversicherungsprämie insgesamt):

  Ihr Gewinn = Ihre Anfangsausstattung - Ihre Kosten für Gesundheitsprävention - Ihre Gruppenversicherungsprämie,

  d. h. Sie zahlen Ihre Kosten für Gesundheitsprävention und Ihre Gruppenversicherungsprämie, währenddessen die Gruppenversicherung im Krankheitsfall Ihre gesamten Behandlungskosten übernimmt.

- Sie sind **individuell versichert** (Zahlungsbereitschaft < Ihre Gruppenversicherungsprämie insgesamt, wenn Sie gruppenversichert wären):

  Ihr Gewinn = Ihre Anfangsausstattung - Ihre Kosten für Gesundheitsprävention - Ihre individuelle Versicherungsprämie,

  d. h. Sie zahlen Ihre Kosten für Gesundheitsprävention und Ihre individuelle Versicherungsprämie, währenddessen die individuelle Versicherung im Krankheitsfall Ihre gesamten Behandlungskosten übernimmt.

- Sie sind **nicht versichert** (Zahlungsbereitschaft < Ihre Gruppenversicherungsprämie insgesamt, wenn Sie gruppenversichert wären):

  - ... und **erkrankt**:

    Ihr Gewinn = Ihre Anfangsausstattung - Ihre Kosten für Gesundheitsprävention - Ihre Behandlungskosten

  - ... und **nicht erkrankt**:

    Ihr Gewinn = Ihre Anfangsausstattung - Ihre Kosten für Gesundheitsprävention,

  d. h. Sie zahlen Ihre Kosten für Gesundheitsprävention und im Krankheitsfall Ihre Behandlungskosten. Eine Versicherungsprämie entfällt.

Prämienberechnung:

- **Individuelle Versicherung:**

  Ihre Prämie = Behandlungskosten × Ihr Gesamtrisiko nach Gesundheitsprävention zu erkranken (in Prozent)/100 ,

  d. h. die individuelle Versicherungsprämie entspricht Ihren erwarteten Behandlungskosten nach Gesundheitsprävention (erwartete Behandlungskosten = Behandlungskosten x aktuelle Wahrscheinlichkeit zu erkranken (in Prozent)/100).

- **Gruppenversicherung:**

  Ihre Prämie insgesamt nach Berücksichtigung der Prämienreduktion durch Gesundheitsprävention = Behandlungskosten × durchschnittliches Gesamtrisiko der Mitglieder der Gruppenversicherung vor Gesundheitsprävention zu erkranken (in Prozent)/100 - Ihre Prämienreduktion durch Gesundheitsprävention

Ihre Prämienreduktion durch Gesundheitsprävention = Behandlungskosten × Ihre Reduktion des Verhaltensrisikos durch Gesundheitsprävention (in Prozentpunkten)/100,

d. h. die Gruppenversicherungsprämie entspricht den erwarteten Behandlungskosten des durchschnittlichen Gruppenversicherten vor Gesundheitsprävention abzüglich den erwarteten Kosteneinsparungen durch Ihre Gesundheitsprävention. Im Gegensatz zur individuellen Versicherungsprämie hängt die Gruppenversicherungsprämie nicht nur von Ihrem genetischen Krankheitsrisiko und Ihrer Gesundheitsprävention ab, sondern auch von den genetischen Krankheitsrisiken der anderen Gruppenversicherten.

Je nach Zahlungsbereitschaften innerhalb einer Gesellschaft kann es mehrere mögliche Gruppenversicherungen in einer Gesellschaft geben, die sich in Bezug auf den Versichertenpool und die Prämie unterscheiden. In diesem Fall wird zunächst die Gruppenversicherung mit der grösstmöglichen Anzahl an Gruppenversicherten gewählt. Sollten dann immer noch mehrere Gruppenversicherungen mit gleicher Versichertenzahl existieren, so wird innerhalb dieser Gruppenversicherungen die Gruppenversicherung mit der grösstmöglichen Anzahl an Gruppenversicherten mit hohem genetischem Risiko zu erkranken gewählt. Sollten dann immer noch mehrere Gruppenversicherungen mit gleicher Versichertenzahl und gleicher Anzahl an Gruppenversicherten mit hohem genetischem Risiko zu erkranken existieren, so wird eine dieser Gruppenversicherungen zufällig ausgewählt.

Gutscheine für den Functional Movement Screen:

Der „Functional Movement Screen“ ist ein standardisiertes Testverfahren aus Amerika, welches zur Erfassung ineffizienter und schädlicher Bewegungsmuster eingesetzt wird. Hauptziel dieses Screens ist es Bewegungsschwächen zu erkennen und Bewegungsabläufe zu verbessern, um so einer Abnutzung und Schädigung des Bewegungsapparates vorzubeugen. Langfristig sind sowohl die Abnutzung als auch die Schädigung des Bewegungsapparates mit starken Schmerzen verbunden und können zu hohen Behandlungskosten führen (z. B. durch die Behandlung bei einem Orthopäden oder Physiotherapeuten).

Der „Functional Movement Screen“ des ASVZs wird durch ausgebildete Physiotherapeuten angeboten. Er umfasst sieben einfache Bewegungstests zur Quantifizierung von Beweglichkeit, Stabilität und Bewegungsmustern. Die Kosten für diesen Screen belaufen sich auf 60 Franken. Der Zeitaufwand für diesen Screen beträgt 30 Minuten. Termine für diesen Screen sind individuell vereinbar. Weitere Informationen zu diesem Präventionsangebot des ASVZs finden Sie auf der Seite des ASVZs oder erhalten Sie unter der Nummer +41 44 251 60 51.

Wenn Sie den Gutschein erhalten (gemäss der Wahrscheinlichkeit des von Ihnen gewählten Levels der Gesundheitsprävention), deckt der Gutschein die gesamten Kosten dieses Screens.

Teil 2

Informationen zu den verschiedenen Situationen, in denen Sie Entscheidungen treffen, erhalten Sie nach Teil 1 des Experiments auf Ihrem Bildschirm.

Instructions for the Experiment

Thank you for your participation in the experiment. Please read the following information carefully. If you have any questions regarding the instructions, please raise your hand. Then, we will come to your cabin to answer your questions. Please do not talk to other participants any longer until the end of the experiment.

For showing up on time, you will receive 10 Swiss Francs. Throughout the experiment, you can earn more money. The amount of your remuneration depends on your decisions and the decisions of other participants. In addition, you have the opportunity to win a voucher for a health preventative measure, which is offered by the student sports association of the university (ASVZ). All decisions are made anonymously, i.e. none of the other participants learns about your identity. Also the final payoff at the end of the experiment is made anonymously, i.e. none of the other participants is informed about your final payoff. Throughout the experiment, the profits are indicated in ECU (= Experimental Currency Unit):

20 ECU = 1 Swiss Franc.

The experiment consists of **two parts**:

- In the **first** part of the experiment, you will make the same sequence of decisions over several periods. At the end of the experiment, **one of these periods is randomly** selected and determines your payoff in this part of the experiment.
- In the **second** part of the experiment, you will face several distinct situations, in which you have to make one or several decisions. At the end of the experiment, **one of these situations is randomly** selected and determines your payoff in this part of the experiment.

After the experiment, we will still ask you to answer some questions. On the following pages, we will explain the exact procedure of the experiment.

Part 1

General Information:

- Part 1 of the experiment consists of 10 periods. In each of these 10 periods, you make the same sequence of decisions.
- In the experiment, in each period you receive an **initial endowment** of 1000 ECU and you face the risk of illness. If you turn ill, this leads to **treatment cost** of 700 ECU.
- Your overall risk of illness is composed of two components: a genetic component and a behavioral component. *In the following, we will simply refer to these components as* ***genetic risk*** *and* ***behavioral risk****.* Your genetic risk, i.e. your innate probability to turn ill, is either low or high. Your genetic risk is randomly assigned to you and it does not change throughout the experiment. Your genetic risk cannot be influenced by you in the experiment. Your initial behavioral risk, i.e. your probability to turn ill because of your behavior, amounts to 20%. In the experiment, your behavioral risk of illness can be influenced by you through the means of health prevention. Your overall risk of illness is the sum of the genetic and behavioral risk. Both components and the resulting overall risk of illness are shown on your computer screen at the beginning of the experiment. This information is private and not observed by other participants of the experiment.

Low risk	
Risk component	Probability to turn ill
Genetic risk	20%
Behavioral risk	20%
Overall risk	40%

High risk	
Risk component	Probability to turn ill
Genetic risk	40%
Behavioral risk	20%
Overall risk	60%

- At the beginning of part 1, you are also randomly assigned to a society. Each society consists of 8 participants, 4 of whom have a high genetic risk of illness and 4 of whom have a low genetic risk of illness. The composition of your society does not change throughout the experiment.
- In each period, you can make two types of decisions: On the one hand, a health prevention decision, which may serve to reduce your behavioral risk of illness, and on the other hand, a health insurance decision.

Sequence of Events in a Period:

1. Each participant receives an initial endowment of 1000 ECU and observes her genetic risk of illness. Each participant chooses among 11 levels how much health prevention she wants to do in order to reduce her behavioral risk of illness. The higher the level of health prevention, the lower the behavioral risk of illness and the higher the probability to obtain a voucher for a health preventative measure of the ASVZ. This measure comprises a "Functional Movement Screen" (FMS). The goal of this screen is to improve the course of motion in order to prevent degeneration and damage of the musculoskeletal system. You will find more information about this screen in the section "Vouchers for the Functional Movement Screen".

   In the following table, you observe the reduction of behavioral risk, the probability to obtain a voucher for the FMS, and the cost of health prevention for each of the 11 corresponding levels of health prevention:

Level of health prevention	Reduction of behavioral risk (in percentage points)	Probability to obtain the voucher for the FMS (in percent)	Cost of health prevention (in ECU)
0	0	0	0
1	2	1	8
2	4	2.5	18
3	6	4.5	30
4	8	7	46
5	10	10	66
6	12	13.5	90
7	14	17.5	118
8	16	22	150
9	18	27	186
10	20	33	226

   That is, if you, for example, have a high genetic risk of illness (40%) and choose the prevention level 6 to reduce your behavioral risk of illness, then your overall risk of illness after health prevention amounts to $40\% + 20\% - 12\% = 48\%$. Due to the health prevention you obtain the voucher for the FMS at the ASVZ with a probability of 13.5%. The cost of your health prevention amounts to 90 ECU.

2. Each participant makes her health insurance choice. In principle, two possibilities for a health insurance exist: either a group insurance or an individual insurance. Both types

of insurance will pay the full treatment cost in the case of illness. If you have the group insurance, then you will be jointly insured with other participants of your society. The insurance premium will then be influenced by you **and** other participants of the group insurance (cf. section "Premium Calculation"). If you have the individual insurance, you will be insured independently of the other members of your society, i.e. in this case the insurance premium will only be influenced by you. You can also remain uninsured. If you are not insured, the treatment cost will be subtracted from your initial endowment in the case of illness.

The health insurance choice in a given period involves two decisions, which are made simultaneously:

(i) Each participant decides on the insurance premium that she is willing to pay at most in order to join the **group insurance**. The group insurance premium is composed of two parts: a part that is identical for all members of the group insurance, and an individual-specific part, which accounts for the premium reduction resulting from health prevention:

- The expected treatment cost of the average group insurance member before health prevention
- **minus** your premium discount due to health prevention.

The willingness to pay determines whether the participant is insured by the group insurance or not. To be group insured at the end of the period, the willingness to pay has to equal at least the total insurance premium that would result if the participant was included in the group insurance. Therefore, it may be that a participant is not group insured, even if her willingness to pay exceeds the total insurance premium for the group insurance. This would be the case, for example, if upon consideration of the participant the total group insurance premium increased to the extent that it lay above the participant's willingness to pay. Each participant can indicate a willingness to pay which equals at most her initial endowment minus her investment in health prevention (according to her chosen level of health prevention).

(ii) For the case that a participant would not be group insured, given her maximum willingness to pay for group insurance, each participant decides whether she wants to insure her risk of illness by an **individual insurance** or whether she wants to remain **uninsured**.

3. The premiums of the group insurance are calculated. A group insurance will be provided if this insurance does not make any losses in expectation, i.e. if the sum of the group insurance premiums is not smaller than the expected treatment costs of all group insurance members. Depending on the willingness to pay in a society, the number of participants who are jointly insured by the group insurance may vary. It may be, for

example, that 3 participants of a society are group insured while the other participants of the society are individually insured or not insured. If no group insurance is provided as the group insurance would make losses in expectation, each participant is either individually insured or not insured according to her decision.

4. For each participant, it is randomly determined whether she turns ill in the period or not. The probability to turn ill corresponds to her actual overall risk of illness, i.e. her overall risk after the reduction of her behavioral risk.

5. At the end of the period, each participant obtains some information about the group insurance: number of members in total, number of members with high genetic risk of illness, insurance premium (before consideration of the premium discount due to health prevention), and her premium discount due to health prevention. Moreover, each participant observes her insurance status (group insured, individually insured, not insured), her insurance premium (after consideration of the premium discount due to health prevention), her illness status (ill, not ill), and her profit.

Your Profit in a Period:

There are four cases, which need to be distinguished:

- You are **group insured** (willingness to pay $\geq$ your total group insurance premium):

  Your profit = your initial endowment - your cost of health prevention - your group insurance premium,

  i.e. you pay your cost of health prevention and your group insurance premium, while the group insurance covers your full treatment cost in the case of illness.

- You are **individually insured** (willingness to pay $<$ your total group insurance premium if you were included in the group insurance):

  Your profit = your initial endowment - your cost of health prevention - your individual insurance premium,

  i.e. you pay the cost of health prevention and your individual insurance premium, while the individual insurance covers your full treatment cost in the case of illness.

- You are **not insured** (willingness to pay < your total group insurance premium if you were included in the group insurance):
  - ... and **ill**:

    Your profit = your initial endowment - your cost of health prevention - your treatment cost
  - ... and **not ill**:

    Your profit = your initial endowment - your cost of health prevention,

  i.e. you pay your cost of health prevention and, in the case of illness, your treatment cost. An insurance premium does not need to be paid.

Premium Calculation:

- **Individual insurance:**

  Your premium = treatment cost × your overall risk of illness after health prevention (in percent)/100,

  i.e. the individual insurance premium corresponds to your expected treatment cost after health prevention (expected treatment cost = treatment cost × actual probability to turn ill (in percent)/100).

- **Group insurance:**

  Your total premium after consideration of the premium discount due to health prevention = treatment cost × average overall risk of illness of all group insurance members before health prevention (in percent)/100 - your premium discount due to health prevention

  Your premium discount due to health prevention = treatment cost × your reduction of behavioral risk resulting from health prevention (in percentage points)/100,

  i.e. the group insurance premium corresponds to the expected treatment cost of the average group insurance member before health prevention minus the expected cost savings due to your health prevention. In contrast to the individual insurance premium, the group insurance premium depends not only on your genetic risk of illness and your health prevention, but also on the genetic risks of illness of the other group insurance members.

Depending on the willingness to pay in a society, there may exist several possible group insurances in a society, which differ with respect to the pool of insured and the premium. In this case, the group insurance with the highest number of group insurance members is selected first. If there still exist several group insurances with the same number of group insurance members, the group insurance with the highest number of group insurance members with high genetic risk is selected among the remaining ones next. If there still exist several group insurances with the same number of group insurance members and the same number of group insurance members with high genetic risk of illness, one of the remaining group insurances is selected at random.

Vouchers for the Functional Movement Screen:

The "Functional Movement Screen" is a standardized test procedure that was developed in the United States and is used to detect inefficient and harmful movement patterns. The primary goal of this screen is that of detecting weaknesses in movement orders and impoving the course of motion in order to prevent degeneration and damage of the musculoskeletal system. In the long run, degeneration as well as damage of the musculoskeletal system causes strong pain and may lead to high treatment costs (e.g., due to the treatment by an orthopedic specialist or a physiotherapist).

The "Functional Movement Screen" at the ASVZ is offered by professionally trained physiotherapists. It comprises seven simple movement tests to quantify mobility, stability, and movement patterns. The cost of this screening amounts to 60 Swiss Francs. The time required for this screening is 30 minutes. Appointments for this screening are made individually. More information about this health preventative measure of the ASVZ is found on the webpage of the ASVZ or is obtained by dialing the number +41 44 251 60 51.

If you receive the voucher (according to the probability that is attached to your chosen level of health prevention), this voucher will cover the full cost of this screening.

Part 2

After completion of part 1 of the experiment, you will receive further information about the distinct decision situations, which you will be facing, on your computer screen.

Curriculum Vitae

Janina Nemitz

Born September 22nd 1988
in Quakenbrück, Germany
German national

Education

08/2013 – 07/2018	Doctoral program at the *Zurich Graduate School of Economics*, University of Zurich, Switzerland
03/2015 – 02/2017	International Doctoral Courses in *Health Economics and Policy*, Swiss School of Public Health, Switzerland
10/2010 – 03/2013	Master of Science in *Quantitative Economics*, University of Konstanz, Germany
08/2011 – 05/2012	Visiting student, Queen's University, Canada
10/2007 – 08/2010	Bachelor of Science in *Economics*, University of Konstanz, Germany

Professional experience

08/2013 – 07/2018	Research and teaching assistant at the Department of Economics, University of Zurich, Switzerland
03/2013 – 07/2013	Internship at the German Institute for Economic Research (DIW), Berlin, Germany
08/2012 – 01/2013	Assistant at the International Office, University of Konstanz, Germany
10/2008 – 09/2011	Research and teaching assistant at the Department of Economics, University of Konstanz, Germany
02/2008 – 03/2008	Internship at the European Parliament, Brussels, Belgium
08/2007 – 09/2007	Internship at DIRRITAS GmbH, Quakenbrück, Germany

March, 2018